Narratives of Collective Action in YouTube's Discourse on Veganism

Arianna Pera,[*] **Luca Maria Aiello**[†]

IT University of Copenhagen
[*]arpe@itu.dk, [†]luai@itu.dk

Abstract

Narratives can be powerful tools for inspiring action on pressing societal issues such as climate change. While social science theories offer frameworks for understanding the narratives that arise within collective movements, these are rarely applied to the vast data available from social media platforms, which play a significant role in shaping public opinion and mobilizing collective action. This gap in the empirical evaluation of online narratives limits our understanding of their relationship with public response. In this study, we focus on plant-based diets as a form of pro-environmental action and employ Natural Language Processing to operationalize a theoretical framework of moral narratives specific to the vegan movement. We apply this framework to narratives found in YouTube videos promoting environmental initiatives such as Veganuary, Meatless March, and No Meat May. Our analysis reveals that several narrative types, as defined by the theory, are empirically present in the data. To identify narratives with the potential to elicit positive public engagement, we used text processing to estimate the proportion of comments supporting collective action across narrative types. Video narratives advocating social fight, whether through protest or through efforts to convert others to the cause, are associated with a stronger sense of collective action in the respective comments. These narrative types also demonstrate increased semantic coherence and alignment between the message and public response, markers typically associated with successful collective action. Our work offers new insights into the complex factors that influence the emergence of collective action, thereby informing the development of effective communication strategies within social movements.

Introduction

Storytelling is a cornerstone of modern communication, with its influence permeating diverse realms from marketing to political discourse (Lund, Cohen, and Scarles 2018; Seargeant 2020). The power of compelling narratives extends beyond individual persuasion and holds the potential to catalyze collective shifts in opinion, mobilize consensus, and foster social cooperation toward shared goals. Social media have democratized opportunities for creating and sharing narratives on a global scale, empowering individuals to shape opinion formation, democratic deliberation, and collective action (Yasseri et al. 2016; Jennings,

Suzuki, and Hubbard 2021; Monti et al. 2022). In particular, video-sharing websites like YouTube have emerged as powerful platforms for disseminating information and mobilizing communities (Uldam and Askanius 2013).

In the past decade, social media facilitated the rapid emergence of social movements that weave narratives in favor of environmental protection and climate action. Research suggests that climate discussions on social platforms, especially those related to initiatives that focus on daily life activities, can significantly enhance public awareness of environmental issues (Mavrodieva et al. 2019). One such activity is the transition to a plant-based diet, which is seen as a tangible and effective action to mitigate the environmental damage caused by over-consumption of animal-based products (Masson-Delmotte et al. 2019; Judge, Fernando, and Begeny 2022). Over the past decade, pro-vegan and pro-vegetarian movements have launched various *challenges* to raise awareness and encourage the public's transition towards a plant-based diet. A prominent example is *Veganuary*, an initiative started in 2014 encouraging people to embrace the vegan diet for the entire month of January each year. In 2023, the initiative attracted over 700k official participants worldwide (veganuary.com 2023). The two other largest initiatives of this kind, spanning over a month each year, are *Meatless March* and *No Meat May*. The success of these challenges relies heavily on participants and activists sharing their experiences online, as this generally fosters a sense of community and increases engagement (Hou 2023).

The effectiveness of these pro-environmental campaigns is thought to be closely linked to the types of narratives that they employ (Fernandez et al. 2015). The framing of environmental issues can influence attention and shape participation in political debates (Nisbet and Newman 2015), for instance impacting the reactions to environmental scandals (Torelli, Balluchi, and Lazzini 2020). Incorporating persuasive communication strategies into pro-environmental messages can shape the public perceptions of climate change and facilitate behavioral shifts (Pelletier and Sharp 2008). However, despite the extensive literature on storytelling in social movements (Davis 2002; Fiskio 2012; Napoli and Ouschan 2020) and the widely reported presence of collective action in climate change discourse (Bamberg, Rees, and Seebauer 2015; Van Zomeren, Pauls, and Cohen-Chen 2019; Schmitt et al. 2019; Hamann and Reese 2020; Furlong and

Vignoles 2021; Gulliver, Fielding, and Louis 2021; Judge, Fernando, and Begeny 2022; Suitner et al. 2023), there is very little quantitative evidence of how much different narratives on climate change covered by theoretical research are effectively used in social media, and how strongly they are associated with a public response that pledges support to climate action. Addressing this gap is crucial to assess the applicability of existing theoretical frameworks to common social media practices and to enhance our understanding of the persuasive potential of different narrative types.

In an effort to bridge this knowledge gap, we delved into an exploration of vegan movements and their social campaigns as a form of pro-environmental action. We operationalized a state-of-the-art theoretical framework of narratives that classifies stories pertinent to the vegan ideology using moral foundations (Napoli and Ouschan 2020). We applied this framework to YouTube videos featuring plant-based transition challenges (such as *Veganuary*, *Meatless March*, and *No Meat May*) to categorize the narratives they present. We did so to answer the following research questions:

RQ1. *How closely do narratives used in YouTube discussions about plant-based diets align with existing theoretical categories?*

RQ2. *Are certain narrative types more frequently associated with a public response advocating for collective action?*

RQ3. *What are the factors that can explain the emergence of collective action responses in the reactions to videos promoting plant-based diets?*

Our study's contribution lies in (i) operationalizing an existing social science theoretical framework of moral narratives used in vegan discourse and mapping YouTube video transcripts to it, (ii) examining the extent to which the language used in video comments elicits the concept of collective action, and (iii) evaluating the significance of various factors in explaining the emergence of collective action language markers in comments.

Our findings indicate that videos whose narratives promote social fight, either in the form of protesting or attempting to convert others to the cause, tended to attract reactions characterized by a more frequent presence of linguistic markers that hint at collective action. These videos, along with those emphasizing freedom of choice, demonstrated the highest levels of semantic consistency across creators, and the highest semantic alignment between the video transcript and the respective comments. Overall, we found that an increased frequency of mentions of collective action in video reactions was significantly associated with a smaller size of the crowd of commenters, a greater content coherence of a video within the reference narrative cluster, and a higher frequency of linguistic markers expressing *loyalty*, a moral dimension that characterizes narratives about social fight, freedom of choice, and duty to educate others through inspirational examples.

Theoretical framework

Social science literature provides various theoretical frameworks for understanding the role of storytelling in the con-text of social movements (see Related Work). While such movements originate from societal changes occurring at a macro scale, their micro-level emergence is shaped by cognitive processes, often triggered by collective action *frames* (Johnston and Noakes 2005). These frames serve as structures that shape individuals' perceptions and motivations within a movement. *Narratives* can be seen as the embodiment of these frames, used by opinion leaders as persuasive tools.

Given our specific interest in plant-based diets as a response to climate change, we narrow our focus to narratives within the vegan movement. Within this narrowed scope, scholars have primarily relied on qualitative studies and interviews to infer common narrative patterns. Through an in-depth analysis of vegan blogs, Napoli and Ouschan (2020) found six narrative types that are described with a combination of moral foundations and identity framing that emerge from the stories. Compared to other theoretical categorizations of vegan narratives types (Waters 2022; Aavik 2023), the framework by Napoli and Ouschan (2020) is semantically more comprehensive, and offers an operationalizable map within the context of online content. Therefore, we adopt such a narrative framework as our theoretical reference, focusing on the vegan movement as a specific case of activism. While the framework predominantly emphasizes animal welfare over explicit climate change concerns, its broad scope allows for a versatile shift towards generic environmental protection without sacrificing relevance. In the remainder of the paper, we will refer to the framework as the *moral vegan ideology*. Next, we will describe the framework itself, outline its theoretical underpinnings, and propose a way to operationalize it.

Veganism narratives

The *moral vegan ideology* framework delineates two groups of archetypal narratives based on the type of collective identity they express: *communal-oriented* and *agency-oriented*. This distinction is based on whether concern for others or self-interest is featured more prominently when presenting the efforts required for driving social change. This partition aligns with the dichotomous moral positioning observed in empirical psychological studies (Frimer and Walker 2009), and with established dimensions of social cognition (Cuddy, Fiske, and Glick 2008). Each of these two groups contains three narrative types that are characterized by distinct moral foundations integral to the vegan ideology: sanctity of life, enacting the authentic self, and freedom. This categorization results in the following narrative types: *duty to educate*, *duty to care*, and *duty to fight* for the communal-oriented identity, and *right to good health*, *right to inner peace*, and *right to choose* for the agency-oriented identity. Table 1 provides a summary of the six narratives and their associated subcategories extracted from the original theoretical formulation of the framework.

The concept of morality intertwines with the distinct narrative types within a specific collective identity group. The *duty to care* and *right to good health* narratives, both linked to the sanctity of life foundation, embody respectively the collective imperative to prioritize animal welfare alongside

	Communal oriented	Agency oriented
Sanctity of life	*Duty to care* (innocence loss, harm to animals, confronting content)	*Right to good health* (**personal** battles, health **benefits**, compassionate relationship)
Enacting the authentic self	*Duty to educate* (**inform**, **inspire** the community, status remark)	*Right to inner peace* (self-acceptance, tension with pre-vegan life)
Freedom	*Duty to fight* (**convert** others, **protest**/boycott, harm to animals)	*Right to choose* (**personal** story, **discuss** diversity of meanings, respect)

Table 1: Six narratives types introduced by Napoli and Ouschan (2020), defined as a combination of identity expression (columns) and moral foundations (rows). Each narrative is divided into sub-categories shown in parenthesis. Sub-categories highlighted in **bold** are those that we found empirically in YouTube videos through our experiments.

that of humans and the value of personal health and well-being. In the pursuit of enacting the authentic self, the *duty to educate* narrative strives to inform and inspire others, particularly through community networks, while the *right to inner peace* narrative works towards aligning actions and values to promote self-acceptance. Finally, narratives rooted in the freedom moral foundation include the *duty to fight*, focusing on converting others to the vegan cause and generating global change for animal protection through protests, product boycotting, and lobbying initiatives, and the *right to choose*, grounded in a philosophy of respect, inclusivity, and kindness.

Operationalization

The *moral vegan ideology* relies on two key concepts for distinguishing narrative types: collective identity and moral foundations. We operationalize them separately, as follows.

Collective identity   To operationalize collective identity, we extracted identity markers from text, focusing on first-person pronouns in both singular and plural forms. Previous studies have demonstrated the predictive nature of these pronouns in expressing collective identity in terms of agency and communion (Decter-Frain and Frimer 2016). We derived a *Collective Identity Index* (CI Index) from pronoun fractions:

$$\text{CI Index} = 0.5 + \frac{0.5 \cdot (f_{\text{I}} - f_{\text{we}})}{f_{\text{I}} + f_{\text{we}} + 1} \qquad (1)$$

where $f_{\bullet}$ denotes the relative frequency of a pronoun group in the text. The index is centered around 0.5: values around 0.5 suggest a balance between the use of first-person singular and plural pronouns, while values closer to 0 or 1 indicate a focus on the community or the self, respectively.

Values shifted toward the two extremes of the range denote "communal-oriented" ($index \leq 0.4$) vs. "agency-oriented" ($index \geq 0.6$) narratives.

Moral foundations   In the realm of moral foundations, the *moral vegan ideology* explicitly acknowledges three crucial factors: *sanctity of life*, *enacting the authentic self*, and *freedom*, which play discriminative roles in defining storytelling types. The Moral Foundation Theory (Graham et al. 2013) (MFT), encompassing pillars such as *care, fairness, loyalty, authority*, and *sanctity*, provides a natural foundation to quantify these factors. Notably, the concepts of *sanctity* and *care* within MFT directly align with *sanctity of life* and *enacting the authentic self* in the narrative framework. Given the inherent relationship between the pillars of MFT and the moral foundations identified in the narrative framework, we opted to leverage MFT as a means to operationalize the latter. While several dictionary-based tools exist for extracting MFT dimensions from text — such as the extensive lexicon developed by (Araque, Gatti, and Kalimeri 2020), which demonstrated good performance on multiple Twitter datasets — they have limitations. Despite their popularity and interpretability, these tools are constrained by fixed and limited vocabularies, only partial handling of word variations, and a tendency for longer documents to have a higher chance of dictionary matches. In response to these limitations, Nguyen et al. (2024) introduced `mformer`, a transformer-based tool for extracting MFT dimensions from text. This tool outputs a number in a range from 0 to 1 for each moral dimension. It has shown superior performance compared to dictionary-based, embedding-based, and supervised classification methods. We employed `mformer` as our reference computational model for extracting moral foundations from MFT. To address potential biases in interpreting standalone moral foundation scores, we implemented a scaling process using a baseline dataset.

Methodological framework

We aim to explore the interplay between narrative types in vegan-related videos and the levels of user engagement in collective action shown in the comments corresponding to these videos. To do so, we devised a three-step computational framework, illustrated in Figure 1. First, we collected and processed YouTube videos and associated comments. Second, we mapped video content to narrative types based on the *moral vegan ideology* framework, thus obtaining distinct narrative clusters. Third and last, we associated these narrative clusters with audience reactions, specifically by investigating the presence of linguistic markers of collective action within video comments. To better interpret the relationship between the narrative clusters and their corresponding comments, we measured how much narratives are clustered in the semantic space, and we analyzed the alignment between the semantics of the video content and that of the corresponding comments.

Data collection

We used the YouTube API to analyze discussions related to three plant-based challenges gathering about 800k offi-
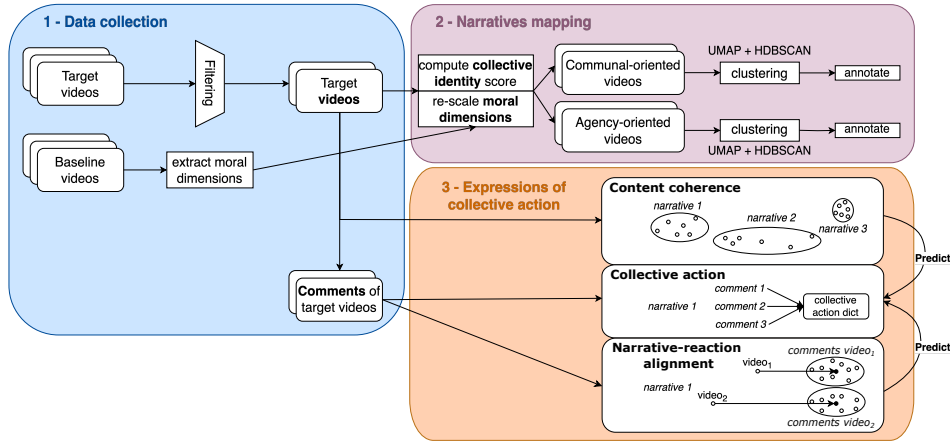
Figure 1: Methodological framework.

	Keywords
Veganuary	*veganuary, #veganuary, #veganuaryYEAR, #Veganuary, #vegan, #Vegan, vegan*
Meatless March	*meatless march, #meatlessmarch, #MeatlessMarch, meatless*
No Meat May	*no meat may, #nomeatmay, #NoMeatMay, vegan, #vegan, #Vegan*

Table 2: Keywords used for the data collection by initiative.

cially registered participants worldwide: *Veganuary*, *Meatless March*, and *No Meat May*. Our selection criteria were driven by the widespread participation and popularity of *Veganuary*, the leading challenge in this domain. In addition, we considered *Meatless March*, and *No Meat May* as they are analogous time-constrained initiatives that enjoy a substantial level of public interest. Our data retrieval spanned videos posted on the platform during the three months centered on the reference month for each challenge, covering a 10-year period from December 2013 to June 2023. We started the collection process by considering the lower-cased names of the movements as a primary keyword (i.e., *veganuary, meatless march, no meat may*) and performed an expansion taking the relevant hashtags that were most often co-occurring with them. In practice, we first queried the API using the seed keywords, extracted the ten most frequent keywords in the descriptions of the set of retrieved videos, and manually selected those that were semantically relevant to the challenge. The final sets of keywords are in Table 2.

We retrieved a total of 12,753 contributions with more than 5M comments. Appendix A provides a yearly breakdown of the data volume for the data of each challenge.

To analyze video content, we leveraged video transcripts. We filtered the set of videos to retain only English content, achieved through the `langdetect` tool for language detection transposed from Java in Python (Shuyo 2010). A subset of videos had the auto-caption option enabled by their creators and thus we were able to retrieve their transcript through the `Youtube Transcript` package (Depoix 2023). To complete the dataset, we employed the `Whisper` audio-to-text converter to obtain the missing data (Radford et al. 2023). This tool has demonstrated performance comparable to humans in terms of Word Error Rate in long-form transcription tasks.

To ensure the selection of pertinent content that could effectively be mapped to theory-defined narratives and evoke specific collective reactions, we employed Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) for topic modeling and conducted a manual inspection of the results. To weed out irrelevant content from the rest, we ran LDA with two as the number of topics, we identified the ten most common words per topic and examined the top-5 most representative video transcripts for each topic. We observed that one topic predominantly revolved around recipes and ingredients, showing no relevance to the *moral vegan ideology* theory, while the other delved into plant-based challenges and their associated ethical aspects. In our narrative mapping approach, the emphasis lies on achieving high precision rather than high recall. This entails prioritizing the accurate identification of pertinent elements, even at the expense of potentially overlooking some less relevant ones. As such, we filtered out recipe-focused videos (i.e. one of the retrieved topics) reaching a total volume of 3,547 data points, 1,801 of which have at least one comment with more than five unique words.

Generally speaking, our pre-processing steps were kept minimal, involving the exclusion of transcripts with fewer than five unique words and the removal of text marked with music tags indicating musical pieces within a YouTube transcript. Table 3 summarizes the filtering steps and resulting data volumes.

For each of the three challenges, we gathered a *baseline set* of videos. These videos served us as a benchmark for re-scaling the metrics calculated on the target video set. To select the baseline, we identified the most frequent YouTube video category for each challenge. Within such categories, we then collected a set of videos of a size similar to the target set (see Appendix A) and within the same reference

Original	English	With transcript	Valid topic	Valid comments
12,753	11,854	8,954 *(YouTube)* / 2,399 *(Whisper)*	3,547	1,801

Table 3: Video data volume by filtering step. Filters are applied in series, left to right.

timeframe, without applying any keyword filter.

Last, we gathered the comments corresponding to our final set of videos. Specifically, we used the `CommentThreads` method in the YouTube API to gather up to 1000 comments and first-level replies for each video. We collected the most recent comments because, at the time of writing, the YouTube API does not allow for retrieving a random sample of comments. In total, we collected 516,207 comments on YouTube, which were further refined to 416,736 after simple pre-processing (removing mentions and URLs, and filtering out comments with fewer than five unique words).

Narratives mapping

To map videos to narrative types defined by the *moral vegan ideology*, we first grouped content by measuring the level of collective identity expressed in the transcripts (Eq. 1). We grouped videos into *communal-oriented* (CI index ≤ 0.4) and *agency-oriented* (CI index ≥ 0.6). We discarded the set of videos with CI index scores between 0.4 and 0.6, as they could not be strongly characterized according to this theoretical dimension. We applied `mformer` to the remaining 3,045 videos to extract their moral foundation scores. To find clusters of narrative types, we used UMAP (McInnes, Healy, and Melville 2018) to reduce the scores to a 2-dimensional space, and finally applied the HDBSCAN clustering algorithm (Campello, Moulavi, and Sander 2013).

We defined optimal HDBSCAN parameters by optimizing the Density Based Cluster Validity (DBCV) (Moulavi et al. 2014) through a random search: we set `min_sample = 15`, `min_cluster_size = 15`, and `metric = manhattan` for the communal-oriented group of videos and `min_sample = 15`, `min_cluster_size = 150`, and `metric = euclidean` for agency-oriented one.

We employed manual annotation to validate the resulting video content clusters. Acknowledging the utility of open-source models like Llama-2 (Touvron et al. 2023) in aiding human annotators in the classification of social science concepts (Ziems et al. 2023), we considered Llama-2 70B Chat as an initial supporting tool. For each cluster in the communal-oriented and agency-oriented groups, we selected the top 30 videos closest to the cluster centroid and prompted the language model to label each transcript within this subset according to the most relevant narrative type, if discernible. We defined narrative types for each identity group in the prompt and instructed it to output "other" in cases of uncertainty (refer to Appendix B for

prompt descriptions). Notably, we observed a tendency of the LLM to predominantly output a single label per collective identity group (e.g., "duty to educate" in the communal-oriented group and "right to good health" in the agency-oriented one), likely influenced by its internal knowledge bias. Given such a first initial guideline, we engaged in manual inspection of transcripts within the top-30 videos subset. This step allowed us to label each cluster and qualitatively define the theme characterizing each narrative group based on sub-categories extracted from the theory, described in Table 1. Data volume for each annotated narrative group can be found in Appendix A.

Collective action expressions

We identified expressions promoting collective action within YouTube comments, aiming to quantify the relationship between these linguistic indicators and various narrative types. This analysis was inspired by theories of normative alignment, which propose that narratives advocating social and political action are crucial in igniting and sustaining collective initiatives (Thomas, McGarty, and Mavor 2009).

To further enrich our understanding of collective action markers, we incorporated two additional semantic similarity measures that previous research has associated with collective action. First, we measured the semantic coherence of videos within a narrative group, as consistent narratives can better foster a shared identity among community members, facilitating collective action efforts (Blair and Jost 2003; Van Zomeren, Postmes, and Spears 2008). Second, we measured the semantic alignment between the content of a video and its corresponding comments as a proxy for agreement between influential leaders and activists, which is typically associated with successful collective movements (Thomas, McGarty, and Mavor 2009). Last, we fit an Ordinary Least Squares (OLS) regression to determine whether the variability in collective action responses to videos is significantly associated with the moral features emerging from them, the semantic coherence with videos' reference narratives, and the video-comments alignment.

Operationalization of collective action Aiming at the definition of a standalone measure of collective action on textual traces, Smith, McGarty, and Thomas (2018) formulated and validated a *collective action dictionary* comprising 47 terms, designed in a LIWC-like fashion (Pennebaker et al. 2015). We used this dictionary to extract collective action instances through a wildcard match of words in comments. We then computed the relative frequency of collective action words in each comment and evaluated the distribution within narrative types.

Content coherence To gauge the semantic consistency of a narrative type, we evaluated the *silhouette score* for each video within it, using S-BERT embeddings (Reimers and Gurevych 2019) (`all-MiniLM-L6-v2` model) of video transcripts. The embedding model, trained on sentence-level data, demonstrates strong performance across various natural language processing downstream tasks. Nonetheless, its effectiveness may fluctuate when handling very large documents, often due to truncation issues. The silhouette score
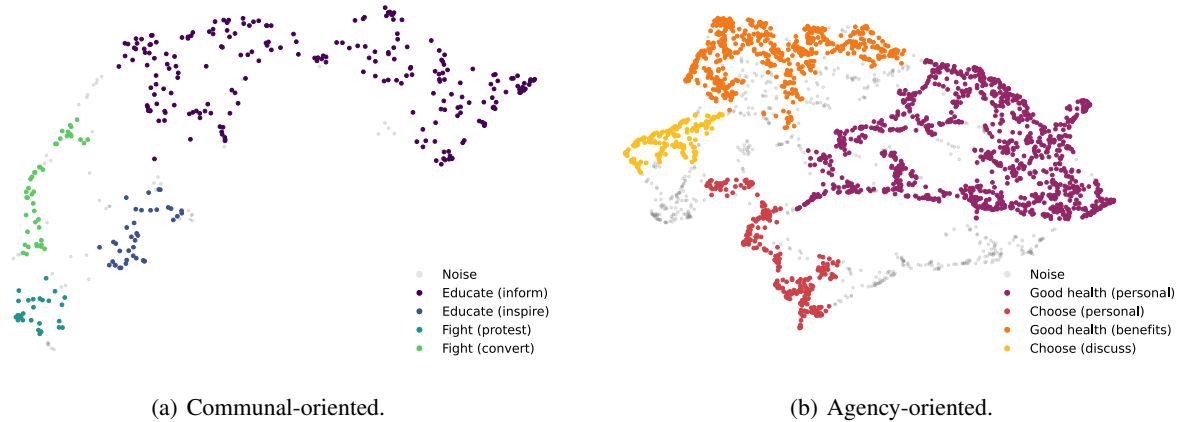
(a) Communal-oriented.

(b) Agency-oriented.

Figure 2: Clusters of videos and their mapping to the *moral vegan ideology* framework.

Narrative	Transcript excerpts
Educate (inform)	Spanish startup creates 3D printed vegan meat. An Italian bioengineer at a Spanish startup claims to have invented the world's first vegan steak made using 3D printing technology...
Educate (inspire)	I'm Matthew Glover, I'm the co-founder of Veganuary... we started it in 2013 we've had four campaigns now. In the first year January 2014 we have 3,300 people taking part and we've been doubling in size or slightly more...
Fight (protest)	...which is why we started this campaign a couple of months ago so um this is an example of a protest we're doing, all of our protests are not quite as destructive as this one...
Fight (convert)	In the afternoon of August 26th 2011, we decided to stop paying other people to kill animals for us to eat... We decided to stop exploiting animals for clothing, entertainment, testing, and all other purposes...
Good health (personal stories)	Being a vegan bodybuilder is really tough. If you're one or want to become one, my hat is off to you because it's really tough...
Choose (personal stories)	I'm not a vegan. Look, man, I'm really proud of you. Cause going vegan is not something that I could do. You make it seem so simple as if I could choose...
Good health (benefits)	Hey this is Ryan of happy healthy vegan. So I just got back results for my latest blood test and I'm gonna share with you right now my testosterone levels...
Choose (discuss)	If you've ever had a conversation with a vegan, especially if the vegan is someone you're close to, you know that things can get heated pretty quickly...

Table 4: Representative examples of YouTube video transcripts for each narrative cluster.

is a widely-used metric to evaluate cluster quality, and measures the similarity of a datapoint with others in its own cluster compared to the similarity with points in other clusters. The distribution of silhouette scores across videos associated with a given narrative type assesses the content coherence of such videos in terms of their placement in the specific narrative cluster.

Narrative-reaction alignment Although our study is solely observational and does not definitively establish a causal link between video content and public reaction, the nature of video-sharing platforms, which are centered around content, provides creators with a degree of influence over emerging topics in public discourse. Thus, we can study the ability of communicators to steer the audience's opinion by analyzing the relationship between published content and related reactions. This relationship can provide insights into the emergence of collective action initiatives within the re-

actions. We considered a *video-comment alignment* metric based on the cosine similarity of pairs of videos and comments S-BERT embeddings. Given $C_i = (c_{i1}, c_{i2}, ..., c_{il})$ comments to video v_i, $\mathbf{v}_i$ S-BERT vector of v_i, and $\overline{\mathbf{C}_i}$ average S-BERT vector of C_i, the metric can be defined as:

$$\textit{video-comment alignment}(v_i) = cos(\overline{\mathbf{C}_i}, \mathbf{v}_i) \quad (2)$$

Results

Narratives

Using the *Collective Identity* index (Eq. 1), we identified 389 communal and 2,656 agency-oriented videos. Within each group, the clustering on the UMAP-reduced dimensions of the Moral Foundations Theory (MFT) resulted in four clusters, as shown in Figure 2. We aligned each cluster with theory-defined narrative types through manual assessment. Table 4 summarizes the labeling of narratives and provides
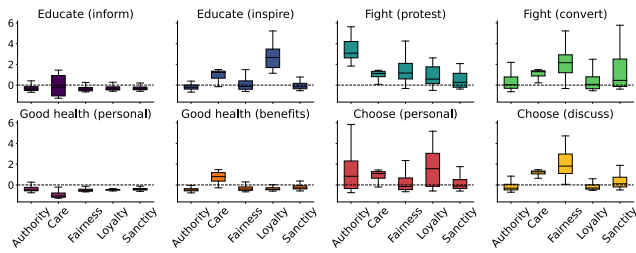
Figure 3: Moral dimensions scores for communal-oriented narratives (top) and agency-oriented narratives (bottom). Scores are discounted by the scores calculated on a control set of videos, and then standardized.

a video transcript example for each of them. We also characterized each narrative based on the distribution of MFT scores of their respective videos (Figure 3).

Within the communal-oriented narratives, two clusters were associated with the *duty to educate* narrative. The first primarily focuses on information sharing, embodying the creator's role as a "conveyor of knowledge" (*educate (inform)*). Videos within this cluster exhibit a marked character of news-like dissemination, therefore their transcripts do not generally show high levels in any MFT dimension. The second cluster emphasizes the inspirational aspect of storytelling, motivating others to start their journey toward veganism or persist in their pursuit (*educate (inspire)*). Its videos exhibit higher levels of loyalty compared to other clusters, reflecting the feeling of "one for all and all for one" typical of that narrative type. The remaining two clusters were linked to the *duty to fight* narrative. The first expresses the intent to persuade others to act towards broader global change (*fight (convert)*), and it is characterized by a high level of the *fairness* MFT dimension, which is linked to the evolutionary process of reciprocal altruism (Brosnan and De Waal 2002). The latter promotes active participation in protests and lobbying to attain climate objectives (*fight (protest)*), and does so by expressing high levels of the *authority* dimension, which stresses the values of leadership and followership.

Within agency-oriented narratives, two clusters refer to concepts typical of *right-to-good-health* narratives: emphasizing personal experiences and transformations (*good health (personal stories)*) and praising the physical and mental health benefits of plant-based diets (*good health (benefits)*). The MFT dimensions for these two clusters are not particularly distinctive, with scores for most dimensions being at the baseline level. The remaining two clusters are semantically associated with the *right-to-choose* narrative, and address the need for broadening the societal acceptance of veganism. The first cluster does so by sharing personal stories (*choose (personal stories)*), which exhibit on average high scores of the *loyalty* dimension being linked to the moral's defining sense of self-sacrifice. The second cluster encourages instead informed discussions (*choose (discuss)*) and the language used in its videos is high on the *fairness* dimension, highlighting the virtues of justice and rights.
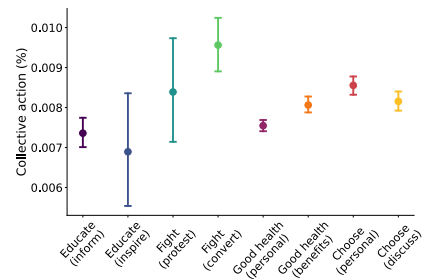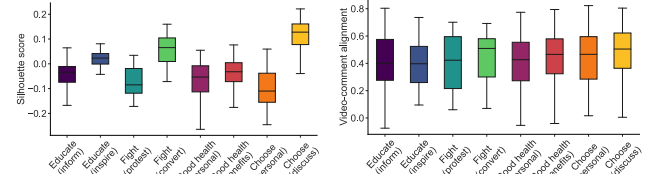

Figure 4: Relative frequency of collective action markers in comments by narrative type.

(a) Silhouette score.   (b) Video-comment align.

Figure 5: Content coherence and narrative-reaction alignment by narrative type.

Collective action in reactions

We examined the frequency of words associated with collective action in comments on videos across narrative groups (Figure 4). We observed a variation in the frequency of these linguistic markers across narratives, with the highest frequency in videos that fell under the *fight (convert)* narrative. Appendix C reports some examples of comments characterized by high levels of collective action.

We further investigated the relationship between the frequency of commitment markers, video content coherence within narrative groups, and narrative-reaction alignment (Eq. 2). The narratives within the *fight (convert)* and *choose (discuss)* clusters demonstrated the highest overall levels of semantic consistency (Figure 5, left). These two narratives were found to occupy a narrow semantic area in the embedding space, as shown by the UMAP projections of the S-BERT embeddings of the video transcripts (Figure 6). We assessed whether the high semantic coherence of certain narrative types translated into a strong semantic alignment between content and reactions (Figure 5). The *fight (convert)* and *choose (discuss)* narratives were characterized by a slightly higher similarity between videos and corresponding comments compared to other clusters. Overall, our findings suggest a correlation between content semantic consistency and the capacity of some narratives to attract reactions aligned with video content.

The observed macro-level association between collective action indicators, narrative types, semantic coherence of content, and alignment of comments to videos, raises the question of whether the frequency of collective action mentions in comments on individual videos can be inferred from these factors. To address this, we employed an Ordinary
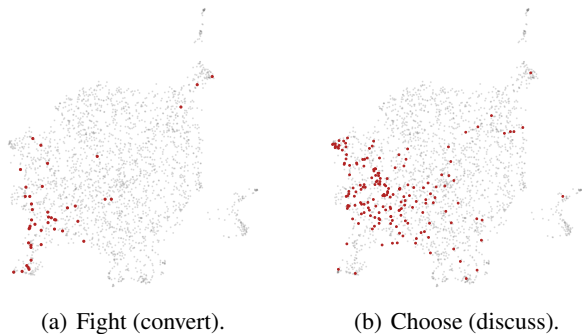
(a) Fight (convert).   (b) Choose (discuss).

Figure 6: UMAP projections of S-BERT video embeddings of all videos, with videos belonging to two narrative types highlighted in red.

Variable	Coefficient	p-value
const	1.16e-17	1.000
n. comments (log)	**-0.059**	**0.025****
video-comment alignment (log)	0.015	0.576
silhouette (log)	**0.047**	**0.066***
care score	-0.034	0.292
fairness score (sqrt)	0.024	0.416
loyalty score (sqrt)	**0.051**	**0.053***
authority score (sqrt)	-0.027	0.329
sanctity score (sqrt)	0.025	0.428
R^2 = 0.008		

Table 5: OLS regression coefficients to predict the fraction of comments containing linguistic markers of collective action.

Least Squares (OLS) model, using individual videos as the unit of analysis, to predict the degree of commitment based on moral foundations scores, semantic coherence between videos and narrative groups, and video-comment alignment. We also factored in the number of comments on each video as a control variable to account for the video's popularity and the content creator's influence. The model was applied after adjusting skewed predictors and calculating the z-scores of all variables, as summarized in Table 5. Despite the inherent challenge in predicting comment commitment variability at the individual video level, we identified several regression coefficients of statistical significance (p-values ≤ 0.1). Specifically, a smaller audience size, higher semantic coherence between content and narrative type, and a discourse more focused on group self-sacrifice (key characteristics of the loyalty moral dimension) were significantly correlated with increased collective action in responses.

Discussion and conclusion

In pursuit of evaluating the influence of online storytelling on collective action initiatives, specifically within the context of climate action, we analyzed YouTube content related to three popular challenges aimed at promoting awareness

about plant-based diets: *Veganuary*, *Meatless March*, and *No Meat May*.

We developed the first operationalization of a theoretical framework that formalizes six moral narratives within the vegan ideology, and mapped YouTube videos to these storytelling types (**RQ1**). For communal-oriented narratives, a clear mapping emerged for the theoretical concepts of "duty to educate" and "duty to fight", while "duty to care" was less distinct in the reference videos. The "duty to educate" narrative frequently surfaced during annotation, likely because of the news-like communication tone of many YouTube vlogs. The distinction between agency-oriented narrative types was more nuanced. The narrative of the "right to good health" was often present, even blending with other narrative types, suggesting a tendency to emphasize the benefits of plant-based diets. The "right to choose" narrative also appeared frequently, while the "right to inner peace" narrative was less common.

Building on the relationship between public narratives and collective action explored in previous theoretical literature, we investigated the connection between narrative types and linguistic markers indicative of collective action in video comments. Narratives advocating for a duty to fight and convert others to the veganism cause elicited the highest level of collective action responses (**RQ2**). Overall, we found that a smaller audience, greater content coherence of videos with their narrative group, and discourse expressing higher levels of the "loyalty" moral foundation were significantly associated with increased collective action (**RQ3**). These findings align with established research in the social sciences. The seminal work of Olson Jr (1971) posited an inverse relationship between the probability of collective action and group size. Moreover, the coherence of content within its designated narrative can cultivate a shared identity among community members, thus fostering more effective collective action efforts (Blair and Jost 2003; Van Zomeren, Postmes, and Spears 2008). Additionally, (Willer 2009) emphasize the intricate relationship between collective action, the benefits accrued within groups (such as status), and the notion of self-sacrifice for the greater collective—a fundamental aspect of the "loyalty" moral foundation. These insights contribute to a nuanced understanding of the dynamics that drive collective action within the context of the proposed analysis.

While our study provides novel quantitative insights into the relationship between narratives and collective action, it has limitations that open up avenues for future research. Primarily, the vegan narrative framework we employed is inherently tailored to its specific topic, as defined within the existing literature. Nonetheless, its structured and clear outline, rooted in collective identity and moral perspectives, presents the opportunity for expansion into other contexts involving lifestyle choices. Other issues arise from the limited amount of data involved in the final steps of the study. Notably, our data pre-processing pipeline retained only 24% of the initial 12,753 videos potentially relevant to the veganism discourse. This was due to several filtering steps aimed at prioritizing precision over recall. Most notably, we excluded video-recipes and all videos not heavily characterized on the Collective Identity axis of our reference theory. Our opera-

tionalization covered only a subset of the narratives originally formulated by the theory, which could be due to these filtering steps or to the limited applicability of the theory to the specific context of YouTube. Another limitation pertains to the complex nature of the concepts characterizing the narratives and the public response they received. Despite using validated tools, the inherent ambiguity and complexity of some of the key theoretical notions we base our work on may have introduced noise in the final mapping between empirical clusters of videos and theoretical narratives. Last, the observed collective action in response to narrative content does not necessarily translate to offline participation in social movements, as social media often fosters a form of activism known as "slacktivism" (Morozov 2009).

Despite these constraints, our study successfully captures the existence of nuances in pro-environmental storytelling and discerns variations in the type of engagement that different narrative types receive.

Related work

Extensive exploration within social science literature has elucidated the profound impact of discourse framing on collective social processes, particularly within the development of social movements. Polletta (1998) identified three dynamics of collective action where stories play pivotal roles, including "stories of origin", "stories of defeat" and "stories of victory". Davis (2002) expanded on this by highlighting the interactive exchange process inherent in narrative frameworks, emphasizing the crucial role of audiences in shaping and responding to stories, with a focus on social identity aspects.

Environmental social movements, the main target of our research, have been often examined through the lens of narrative networks, emphasizing the connection between stories and the social networks formed in the context of environmental protection (Lejano, Ingram, and Ingram 2013). Within this setting, climate change-specific narratives have been categorized into "lifeboat" and "collective" types, reflecting dichotomies between self-interest and shared purpose within communities (Fiskio 2012). The theoretical framework by Napoli and Ouschan (2020), a key reference for our work, further enriched this collective identity-based description of environmental narratives by focusing on the vegan movement and adding nuances of morality. This perspective has been further corroborated by later experimental work (Waters 2022; Aavik 2023).

While numerous contributions exist within the field of Narrative Analysis that extract narratives from textual traces using Natural Language Processing (NLP) tools (Bandeli, Hussain, and Agarwal 2020; Debnath et al. 2020; Ranade et al. 2022), the link between computational methods and narrative types discussed in social science literature is still under-explored (Piper, So, and Bamman 2021). Notable exceptions include the work of Vu et al. (2021), who extracted framing ingredients of climate change-related protest messages, operationalizing concepts such as impacts, action, and efficacy (Benford and Snow 2000) through mention counts.

Shifting the focus to the operationalization of collective action, the literature presents a richer landscape, often examined through network science lenses (Lucchini et al. 2022). Approaches involving textual analysis, particularly the use of recurrent expressions (Berlin 2008; Fetzer 2008), present challenges especially when applied to social media comments. Research has also explored machine learning tools to extract linguistic markers hinting at collective action from text (Wang, Cutler, and Culotta 2017; Zhang and Pan 2019; Li et al. 2023), yet the domain-specific nature of these approaches limits their general applicability. In the realm of dictionary-based methods for collective action measurement, studies have commonly focused on collective identity, emotions, and collective or participatory efficacy as predictors (Van Zomeren, Postmes, and Spears 2008; Van Zomeren, Saguy, and Schellhaas 2013; Rand, Kraft-Todd, and Gruber 2015; Bamberg, Rees, and Seebauer 2015; Van Zomeren, Pauls, and Cohen-Chen 2019; Hamann and Reese 2020; Furlong and Vignoles 2021; Brown, Lowery, and Smith 2022; Judge, Fernando, and Begeny 2022; Thomas et al. 2022). Some have aimed to define computational measures for these elements from text, often resorting to LIWC categories (Gulliver, Fielding, and Louis 2021; Suitner et al. 2023). Within this group of works, Smith, McGarty, and Thomas (2018) introduced a collective action dictionary, defined in a LIWC-like fashion, which served as a valuable resource for our analysis.

Ethical considerations

Given that YouTube content is predominantly user-generated and largely unmoderated, the narratives we examined may contain messages that do not necessarily advocate for veganism and climate action in an ethical or respectful manner. Our study adopted a broad view of the public discourse on YouTube, without attempting to categorize content that may contain misinformation or potentially harmful expressions. We underscore the importance of addressing the potential ethical implications associated with unregulated online spaces.

We recognize the potential misuse of our framework, which could be exploited to inform strategies of misinformation and propaganda. Despite acknowledging the risks of certain aspects being leveraged to create tools and strategies with detrimental societal implications, we are dedicated to minimizing such risks. Our motivation is grounded in the belief that our work can positively contribute to addressing pressing social dilemmas, particularly those related to climate change.

The dataset associated with this paper complies with the FAIR principles (FORCE11 2020), as outlined in the DataSheet (Supplementary Material).

Code and data availability

All code used for the experiments and information about the videos examined, in terms of narrative orientations, statistics, metrics specific to this paper, reference baseline videos, and comments retrieved and analyzed is available on a GitHub repository[1].

---

[1]https://github.com/ariannap13/VeganCollectiveAction

Acknowledgments

References

Aavik, K. 2023. Going Vegan: Understanding Men's Vegan Transition Narratives. In *Contesting Anthropocentric Masculinities Through Veganism: Lived Experiences of Vegan Men*, 33–68. Springer.

Araque, O.; Gatti, L.; and Kalimeri, K. 2020. MoralStrength: Exploiting a moral lexicon and embedding similarity for moral foundations prediction. *Knowledge-based systems*, 191: 105184.

Bamberg, S.; Rees, J.; and Seebauer, S. 2015. Collective climate action: Determinants of participation intention in community-based pro-environmental initiatives. *Journal of Environmental Psychology*, 43: 155–165.

Bandeli, K. K.; Hussain, M. N.; and Agarwal, N. 2020. A Framework towards Computational Narrative Analysis on Blogs. In *Text2Story@ ECIR*, 63–69.

Benford, R. D.; and Snow, D. A. 2000. Framing processes and social movements: An overview and assessment. *Annual review of sociology*, 26(1): 611–639.

Berlin, L. N. 2008. "I Think, Therefore..." Commitment in Political Testimony. *Journal of Language and Social Psychology*, 27(4): 372–383.

Blair, I. V.; and Jost, J. T. 2003. Exit, loyalty, and collective action among workers in a simulated business environment: Interactive effects of group identification and boundary permeability. *Social Justice Research*, 16: 95–108.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan): 993–1022.

Brosnan, S. F.; and De Waal, F. B. 2002. A proximate perspective on reciprocal altruism. *Human Nature*, 13: 129–152.

Brown, O.; Lowery, C.; and Smith, L. G. 2022. How opposing ideological groups use online interactions to justify and mobilise collective action. *European Journal of Social Psychology*, 52(7): 1082–1110.

Campello, R. J.; Moulavi, D.; and Sander, J. 2013. Densitybased clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, 160–172. Springer.

Cuddy, A. J.; Fiske, S. T.; and Glick, P. 2008. Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology*, 40: 61–149.

Davis, J. E. 2002. Narrative and social movements. *Stories of change: Narrative and social movements*, 3–29.

Debnath, R.; Darby, S.; Bardhan, R.; Mohaddes, K.; and Sunikka-Blank, M. 2020. Grounded reality meets machine learning: A deep-narrative analysis framework for energy policy research. *Energy research & social science*, 69: 101704.

Decter-Frain, A.; and Frimer, J. A. 2016. Impressive words: linguistic predictors of public approval of the US congress. *Frontiers in psychology*, 7: 240.

Depoix, J. 2023. Youtube Transcript API. https://github.com/jdepoix/youtube-transcript-api.

Fernandez, M.; Burel, G.; Alani, H.; Piccolo, L. S. G.; Meili, C.; and Hess, R. 2015. Analysing engagement towards the 2014 earth hour campaign in twitter.

Fetzer, A. 2008. "And I Think That Is a Very Straightforward Way of Dealing With It" The Communicative Function of Cognitive Verbs in Political Discourse. *Journal of Language and Social Psychology*, 27(4): 384–396.

Fiskio, J. 2012. Apocalypse and ecotopia: narratives in global climate change discourse. *Race, Gender & Class*, 12–36.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Frimer, J. A.; and Walker, L. J. 2009. Reconciling the self and morality: An empirical model of moral centrality development. *Developmental Psychology*, 45(6): 1669.

Furlong, C.; and Vignoles, V. L. 2021. Social identification in collective climate activism: Predicting participation in the environmental movement, extinction rebellion. *Identity*, 21(1): 20–35.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Graham, J.; Haidt, J.; Koleva, S.; Motyl, M.; Iyer, R.; Wojcik, S. P.; and Ditto, P. H. 2013. Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*, volume 47, 55–130. Elsevier.

Gulliver, R.; Fielding, K. S.; and Louis, W. R. 2021. Assessing the mobilization potential of environmental advocacy communication. *Journal of Environmental Psychology*, 74: 101563.

Hamann, K. R.; and Reese, G. 2020. My influence on the world (of others): Goal efficacy beliefs and efficacy affect predict private, public, and activist pro-environmental behavior. *Journal of Social Issues*, 76(1): 35–53.

Hou, J. Z. 2023. "Sharing Is Caring": Participatory Storytelling and Community Building on Social Media Amidst the COVID-19 Pandemic. *American Behavioral Scientist*, 00027642231164040.

Jennings, F. J.; Suzuki, V. P.; and Hubbard, A. 2021. Social media and democracy: Fostering political deliberation and participation. *Western Journal of Communication*, 85(2): 147–167.

Johnston, H.; and Noakes, J. A. 2005. *Frames of protest: Social movements and the framing perspective*. Rowman & Littlefield Publishers.

Judge, M.; Fernando, J. W.; and Begeny, C. T. 2022. Dietary behaviour as a form of collective action: A social identity model of vegan activism. *Appetite*, 168: 105730.

Lejano, R.; Ingram, M.; and Ingram, H. 2013. *The power of narrative in environmental networks*. MIT Press.

Li, Y.; Frans, V. F.; Song, Y.; Cai, M.; Zhang, Y.; and Liu, J. 2023. SDGdetector: an R-based text mining tool for quantifying efforts toward Sustainable Development Goals. *Journal of Open Source Software*, 8(84): 5124.

Lucchini, L.; Aiello, L. M.; Alessandretti, L.; De Francisci Morales, G.; Starnini, M.; and Baronchelli, A. 2022. From Reddit to Wall Street: The role of committed minorities in financial collective action. *Royal Society Open Science*, 9(4): 211488.

Lund, N. F.; Cohen, S. A.; and Scarles, C. 2018. The power of social media storytelling in destination branding. *Journal of destination marketing & management*, 8: 271–280.

Masson-Delmotte, V.; Pörtner, H.-O.; Skea, J.; Zhai, P.; Roberts, D.; Shukla, P. R.; Buendía, E. C.; et al. 2019. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Mavrodieva, A. V.; Rachman, O. K.; Harahap, V. B.; and Shaw, R. 2019. Role of social media as a soft power tool in raising public awareness and engagement in addressing climate change. *Climate*, 7(10): 122.

McInnes, L.; Healy, J.; and Melville, J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Monti, C.; Aiello, L. M.; De Francisci Morales, G.; and Bonchi, F. 2022. The language of opinion change on social media under the lens of communicative action. *Scientific Reports*, 12(1): 17920.

Morozov, E. 2009. The brave new world of slacktivism. *Foreign policy*, 19(05).

Moulavi, D.; Jaskowiak, P. A.; Campello, R. J.; Zimek, A.; and Sander, J. 2014. Density-based clustering validation. In *Proceedings of the 2014 SIAM international conference on data mining*, 839–847. SIAM.

Napoli, J.; and Ouschan, R. 2020. Vegan stories: Revealing archetypes and their moral foundations. *Qualitative Market Research: An International Journal*, 23(1): 145–169.

Nguyen, T. D.; Chen, Z.; Carroll, N. G.; Tran, A.; Klein, C.; and Xie, L. 2024. Measuring Moral Dimensions in Social Media with Mformer. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Nisbet, M. C.; and Newman, T. P. 2015. Framing, the media, and environmental communication. In *The Routledge handbook of environment and communication*, 345–358. Routledge.

Olson Jr, M. 1971. *The Logic of Collective Action: Public Goods and the Theory of Groups, with a new preface and appendix*, volume 124. Harvard University Press.

Pelletier, L. G.; and Sharp, E. 2008. Persuasive communication and proenvironmental behaviours: how message tailoring and message framing can improve the integration of behaviours through self-determined motivation. *Canadian Psychology/Psychologie Canadienne*, 49(3): 210.

Pennebaker, J.; Booth, R.; Boyd, R.; and Francis, M. 2015. Linguistic Inquiry and Word Count: LIWC2015.

Piper, A.; So, R. J.; and Bamman, D. 2021. Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 298–311.

Polletta, F. 1998. Contending stories: Narrative in social movements. *Qualitative sociology*, 21: 419–446.

Radford, A.; Kim, J. W.; Xu, T.; Brockman, G.; McLeavey, C.; and Sutskever, I. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, 28492–28518. PMLR.

Ranade, P.; Dey, S.; Joshi, A.; and Finin, T. 2022. Computational understanding of narratives: A survey. *IEEE Access*, 10: 101575–101594.

Rand, D. G.; Kraft-Todd, G.; and Gruber, J. 2015. The collective benefits of feeling good and letting go: Positive emotion and (dis) inhibition interact to predict cooperative behavior. *PloS one*, 10(1): e0117426.

Reimers, N.; and Gurevych, I. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Schmitt, M. T.; Mackay, C. M.; Droogendyk, L. M.; and Payne, D. 2019. What predicts environmental activism? The roles of identification with nature and politicized environmental identity. *Journal of Environmental Psychology*, 61: 20–29.

Seargeant, P. 2020. *The art of political storytelling: Why stories win votes in post-truth politics*. Bloomsbury Publishing.

Shuyo, N. 2010. Language Detection Library for Java.

Smith, L. G.; McGarty, C.; and Thomas, E. F. 2018. After Aylan Kurdi: How tweeting about death, threat, and harm predict increased expressions of solidarity with refugees over time. *Psychological science*, 29(4): 623–634.

Suitner, C.; Badia, L.; Clementel, D.; Iacovissi, L.; Migliorini, M.; Casara, B. G. S.; Solimini, D.; Formanowicz, M.; and Erseghe, T. 2023. The rise of# climateaction in the time of the FridaysForFuture movement: A semantic network analysis. *Social Networks*, 75: 170–185.

Thomas, E. F.; Duncan, L.; McGarty, C.; Louis, W. R.; and Smith, L. G. 2022. MOBILISE: a higher-order integration of collective action research to address global challenges. *Political Psychology*, 43: 107–164.

Thomas, E. F.; McGarty, C.; and Mavor, K. I. 2009. Aligning identities, emotions, and beliefs to create commitment to sustainable social and political action. *Personality and social psychology review*, 13(3): 194–218.

Torelli, R.; Balluchi, F.; and Lazzini, A. 2020. Greenwashing and environmental communication: Effects on stakeholders' perceptions. *Business strategy and the Environment*, 29(2): 407–421.

Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Uldam, J.; and Askanius, T. 2013. Online civic cultures? Debating climate change activism on YouTube. *International Journal of Communication*, 7.

Van Zomeren, M.; Pauls, I. L.; and Cohen-Chen, S. 2019. Is hope good for motivating collective action in the context of climate change? Differentiating hope's emotion-and problem-focused coping functions. *Global Environmental Change*, 58: 101915.

Van Zomeren, M.; Postmes, T.; and Spears, R. 2008. Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. *Psychological bulletin*, 134(4): 504.

Van Zomeren, M.; Saguy, T.; and Schellhaas, F. M. 2013. Believing in "making a difference" to collective efforts: Participative efficacy beliefs as a unique predictor of collective action. *Group Processes & Intergroup Relations*, 16(5): 618–634.

veganuary.com. 2023. Veganuary 2023 Participants Survey.

Vu, H. T.; Blomberg, M.; Seo, H.; Liu, Y.; Shayesteh, F.; and Do, H. V. 2021. Social media and environmental activism: Framing climate change on Facebook by global NGOs. *Science communication*, 43(1): 91–115.

Wang, Z.; Cutler, J.; and Culotta, A. 2017. Are words commensurate with actions? quantifying commitment to a cause from online public messaging. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1050–1057. IEEE.

Waters, C. 2022. To V or Not to V: Contingencies of Going Vegan. *Humanity & Society*, 46(2): 151–169.

Willer, R. 2009. Groups reward individual sacrifice: The status solution to the collective action problem. *American Sociological Review*, 74(1): 23–43.

Yasseri, T.; Margetts, H.; John, P.; and Hale, S. 2016. *Political turbulence: How social media shape collective action*. Princeton University Press.

Zhang, H.; and Pan, J. 2019. Casm: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology*, 49(1): 1–57.

Ziems, C.; Shaikh, O.; Zhang, Z.; Held, W.; Chen, J.; and Yang, D. 2023. Can large language models transform computational social science? *Computational Linguistics*, 1–53.

Appendix A

Table 6 summarizes the volume of collected data in terms of the number of videos and comments per year, for each of the analyzed plant-based challenges. Note that such a volume refers to the originally collected data, i.e. no filtering or cleaning is performed in this phase.

Table 6: Volume of originally retrieved videos and comments by movement and year: Veganuary (v), Meatless March (m), No Meat May (n).

Year	n. videos	n. comments
2014	v: 300 m: 64 n: 314	v: 27113 m: 281 n: 22380
2015	v: 431 m: 91 n: 423	v: 61047 m: 2379 n: 143316
2016	v: 445 m: 152 n: 457	v: 74136 m: 2715 n: 303835
2017	v: 527 m: 175 n: 438	v: 202943 m: 3688 n: 166767
2018	v: 678 m: 207 n: 331	v: 144711 m: 16217 n: 440456
2019	v: 728 m: 240 n: 437	v: 146233 m: 11587 n: 424704
2020	v: 820 m: 336 n: 442	v: 241289 m: 128576 n: 342716
2021	v: 786 m: 392 n: 441	v: 418376 m: 18375 n: 398727
2022	v: 787 m: 380 n: 305	v: 151607 m: 12584 n: 176818
2023	v: 724 m: 385 n: 517	v: 460404 m: 37384 n: 484054

Table 7 compares the volume of the cleaned and filtered collected target data with the volume of baseline videos of reference, for each of the analyzed plant-based challenges.

Figure 7 outlines the distribution of videos by mapped narrative type.

Table 7: Volume of cleaned videos by movement and year, comparing target and baseline sets: Veganuary (v), Meatless March (m), No Meat May (n).

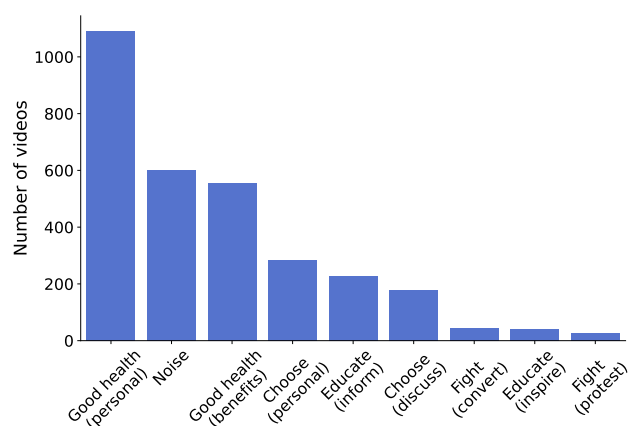| Year	n. videos target	n. videos baseline
2014	v: 78 m: 24 n: 81	v: 76 m: 24 n: 76
2015	v: 133 m: 34 n: 136	v: 129 m: 34 n: 136
2016	v: 160 m: 39 n: 145	v: 141 m: 39 n: 143
2017	v: 182 m: 43 n: 106	v: 173 m: 42 n: 103
2018	v: 237 m: 46 n: 78	v: 220 m: 45 n: 77
2019	v: 291 m: 82 n: 132	v: 233 m: 76 n: 129
2020	v: 290 m: 69 n: 77	v: 246 m: 69 n: 77
2021	v: 232 m: 71 n: 100	v: 219 m: 66 n: 96
2022	v: 166 m: 63 n: 76	v: 151 m: 61 n: 75
2023	v: 173 m: 56 n: 147	v: 125 m: 49 n: 88

Figure 7: Number of videos by narrative type, according to the reference moral vegan ideology framework.

Appendix B

We prompted Llama 2 70B Chat by providing the model with a system prompt to contextualize the task and a hu-

man message to precisely define the task. For each of the narrative identity groups (i.e. communal-oriented and agency-oriented), we included definitions of narratives in the prompts as follows:

- *Duty to care*: Moral duty to protect all living creatures, fostering compassion and connection with all beings.

- *Duty to fight*: Moral duty to dispel ignorance, overcome misinformation, and inspire others about the benefits of a vegan lifestyle.

- *Duty to educate*: Moral duty to actively engage in rights activism, driven by a sense of injustice.

- *Right to good health*: Moral right to prioritize personal well-being through a plant-based diet, emphasizing physical and mental benefits.

- *Right to inner peace*: Moral right to align actions with values, fostering a life of peace and self-acceptance.

- *Right to choose*: Moral right to make personal choices, accepting others' opinion and not imposing own choices.

Narratives - communal-oriented

```
1   System prompt: You are tasked with
        identifying the predominant moral
        narrative expressed by a text.
2
3   Prompt: Identify the predominant
        narrative expressed by the text,
        choosing between 'duty_to_care'
        ({dutycare}), 'duty_to_fight' ({
        dutyfight}), 'duty_to_educate' ({
        dutyeducate}), or 'other' if you
        cannot find an answer. DO NOT
        reply using a complete sentence,
        and ONLY give the answer in the
        following format: LABEL.
4
5   Text: {text}
6
7   Answer:
```

Narratives - agency-oriented

```
1   System prompt: You are tasked with
        identifying the predominant moral
        narrative expressed by a text.
2
3   Prompt: Identify the predominant
        narrative expressed by the text,
        choosing between '
        right_to_good_health' ({
        rightogoodhealth}), '
        right_to_inner_peace' ({
        rightoinnerpeace}), '
        right_to_choose' ({rightochoose})
        , or 'other' if cannot find an
        answer. DO NOT reply using a
        complete sentence, and ONLY give
        the answer in the following
        format: LABEL.
4
5   Text: {text}
6
7   Answer:
```

Appendix C

The following list contains selected examples within the top-50 comments with the highest relative frequency of collective action markers overall:

- "Vegan don't fight, vegans unite."
- "Do you support factory farming?"
- "Do you even fight bro."
- "Don't do over action ok?"
- "So what do we do?"
- "Okay, what do I do?"
- "I support this calm protests."
- "Facebook: you're supporting actual genocide by supporting facebook."
- "Actually we do this everyday here."
- "What do I do about that?"
- "How do you join random meetings?"
- "So, if this guy decides to actively join the cause and become vegan, will you actively join and support his cause?"
- "Join us eaters."
- "How can I do what you do?"

Appendix D

In this section, we will detail existing assets used in terms of their licenses and the computational resources needed for the analysis.

Licenses

Code snippets provided by Google Developers for the use of YouTube API are licensed under the Apache 2.0 License. Python packages used to process YouTube data, namely Youtube Transcript, langdetect

and `Whisper` are licensed under an MIT license, while `Llama-2-70B-chat` model employed in the mapping validation is an open-source LLM released under a commercial use license[2]. `Mformer` models used in the extraction of moral scores are shared on HuggingFace and licensed under an MIT License, while `S-BERT` model for the extraction of sentence embeddings is licensed under the Apache License 2.0. The collective action dictionary by Smith, McGarty, and Thomas (2018) is publicly available.

Resources

We utilized a V100 30GB GPU for both extracting moral foundations scores through `Mformer` and generating sentence embeddings via `S-BERT`. All other resources were deployed locally on an Apple M1 Pro machine with 8 cores and 16GB of RAM.

---

[2]https://ai.meta.com/llama/license/

Paper Checklist

1. For most authors...

   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes.

   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes.

   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, in the "Methodological framework" section.

   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, we discuss potential biases of our data in the "Discussion and conclusion" section.

   (e) Did you describe the limitations of your work? Yes, limitations are presented and discussed in the "Discussion and conclusion" section.

   (f) Did you discuss any potential negative societal impacts of your work? Yes, we address negative societal impact in the "Ethical considerations" section.

   (g) Did you discuss any potential misuse of your work? Yes, we discuss potential misuse in the "'Ethical considerations" section.

   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, we share our framework in the "Methodological framework" section, and full experimental details (code) are provided in the Supplementary Material. Moreover, data does not contain user-specific information, only aggregated or anonymized measures (see DataSheet).

   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.

2. Additionally, if your study involves hypotheses testing...

   (a) Did you clearly state the assumptions underlying all theoretical results? Yes, we frame our experiments within the scope of previous research in the "Methodological framework" section.

   (b) Have you provided justifications for all theoretical results? Yes, justifications are provided in the "Results" and "Discussion and conclusion" sections.

   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes, other approaches in terms of theoretical frameworks and the motivation behind our choice are described in the "Methodological framework" and the "Related works" sections.

   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, in the "Results" and "Discussion and conclusion" sections we explore potential confounders of our results and highlight limitations of our work.

   (e) Did you address potential biases or limitations in your theoretical framework? Yes, these are addressed in the "Discussion and conclusion" section.

   (f) Have you related your theoretical results to the existing literature in social science? Yes, we directly relate our work to previous work in Computational Social Science and Social Psychology theories, as detailed in "Introduction", "Theoretical framework", "Methodological framework" and "Related works".

   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes, these are discussed in the "Discussion and conclusion" section.

3. Additionally, if you are including theoretical proofs...

   (a) Did you state the full set of assumptions of all theoretical results? NA.

   (b) Did you include complete proofs of all theoretical results? NA.

4. Additionally, if you ran machine learning experiments...

   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes. Included link to GitHub repository.

   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, whenever necessary (e.g. HDBSCAN clustering in the narratives mapping and OLS regression predictors). Described in "Methodological framework" and in "Results", details are given in the code provided in the GitHub repository. In all other cases, we do not train any model(s) but use existing ones.

   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes, we report error bars in all relevant Figures (3, 4 and 5).

   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes, see Appendix D.

   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, specified in the "Methodological framework" section.

   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

   (a) If your work uses existing assets, did you cite the creators? Yes, creators of models used are cited in the "Theoretical framework", "Methodological framework" and "Results" sections.

   (b) Did you mention the license of the assets? Yes, see Appendix D.

(c) Did you include any new assets in the supplemental material or as a URL? Yes, generated data is included in the GitHub repository.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes. Our data does not include PII (except for identifiable YouTube channels). We discuss misinformation and offensive content in the "Ethical considerations" section.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? Yes, see 'Ethical considerations" section and DataSheet on the GitHub repository.

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? Yes, see the GitHub repository.

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA.

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA.

(d) Did you discuss how data is stored, shared, and deidentified? NA.