Affect in science communication: a data-driven analysis of TED Talks on YouTube

Fischer, Olivia ; Jeitziner, Loris T ; Wulff, Dirk U

Affect in science communication: a data-driven analysis of TED Talks on YouTube

Olivia Fischer [1,5], Loris T. Jeitziner[2,3,5] & Dirk U. Wulff[3,4✉]

Science communication is evolving: Increasingly, it is directed at the public rather than academic peers. Understanding the circumstances under which the public engages with scientific content is therefore crucial to improving science communication. In this article, we investigate the role of affect on audience engagement with a modern form of science communication: TED Talks on the social media platform YouTube. We examined how two aspects of affect, valence and density are associated with public engagement with the talk in terms of popularity (reflecting views and likes) and polarity (reflecting dislikes and comments). We found that the valence of TED Talks was associated with both popularity and polarity: Positive valence was linked to higher talk popularity and lower talk polarity. Density, on the other hand, was only associated with popularity: Higher affective density was linked to higher popularity—even more so than valence—but not polarity. Moreover, the association between affect and engagement was moderated by talk topic, but not by whether the talk included scientific content. Our results establish affect as an important covariate of audience engagement with scientific content on social media, which science communicators may be able to leverage to steer engagement and increase reach.

[1] University of Zurich, Zurich, Switzerland. [2] University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland. [3] University of Basel, Basel, Switzerland. [4] Max Planck Institute for Human Development, Berlin, Germany. [5] These authors contributed equally: Olivia Fischer, Loris T. Jeitziner. ✉email: dirk.wulff@gmail.com

Introduction

The digital age presents both opportunities for and challenges to science communication. Communication hubs such as Twitter, Facebook, and YouTube offer unprecedented reach for scientific content and interaction with the public (Collins et al. 2016), thereby making science more accessible for scientists and laypeople alike. With engagement tools such as likes, dislikes, comments, and shares, members of the general public now no longer simply consume scientific content but can also disseminate it. As a result, scientific content that does not engage the public may never reach a large audience. In the oversaturated and highly competitive environment of social media, how can scientists make their voices heard?

Science communication via social media differs in at least two important respects from traditional peer-to-peer science communication. First, because social media users tend to consume content more superficially (Boczkowski et al. 2017), surface-level aspects of content such as choice of language are likely more important for gaining a competitive advantage. Second, content on social media can be shared indirectly, through recommender systems (Covington et al. 2016), as well as directly. These differences introduce strong positive feedback between user engagements, which can greatly amplify the reach of highly engaging content (Aldous et al. 2019; Davidson et al. 2010; Hoiles et al. 2017). This means that scientists rely on laypeople to propagate their messages on social media, which in turn incentivizes scientists to pay attention to the aspects of science communication that make it more engaging.

In this article, we investigate affect as one aspect of science communication that may be instrumental for effective science communication (Milkman and Berger, 2014). Past work has found that *New York Times* articles using more affect-rich language were more likely to make the *New York Times* most-emailed list (e.g., Berger and Milkman, 2012). There is also evidence that scientific findings described in a more affective manner are more likely to be shared (Milkman and Berger, 2014) and tend to garner more citations (Fronzetti Colladon et al. 2020). However, the potential link between affect and engagement as a driver of dissemination has not been systematically investigated for social media-based science communication (see Davies, 2019; Davies et al. 2019; Osseweijer, 2006). We aim to fill this gap with a data-driven analysis of engagement with TED Talks on the social media platform YouTube.

TED Talks are short recorded presentations on technology, entertainment, and design; many address basic and applied science. TED Talks are therefore studied as a modern form of science communication (e.g., Gheorghiu et al. 2020; MacKrill et al. 2021; Sugimoto and Thelwall, 2013; Verjovsky and Jurberg, 2020). The transcripts of all talks featured on the TED website (www.ted.com) can be used to derive their affective features. TED Talks are shared on the TED website and on the organization's YouTube channel, which has a total of 19.8 million subscribers and over two billion video views[1]. The popularity of TED Talks on YouTube reflects that they are targeted at a lay audience and contain less jargon (Rakedzon et al. 2017; Sharon and Baram-Tsabari, 2014); these talks therefore offer a rich data trove on public engagement that can be linked to the talks' affective features.

There is a growing body of work on social media-based science communication (see Allgaier, 2020; Brossard, 2013; Kohler and Dietrich, 2021) and, in particular, science communication on YouTube. Past work has focused on understanding the role of characteristics of video presenters for user engagement, including their gender (Amarasekara and Grant, 2019), professional background, and perceived authenticity (Kaul et al. 2020), as well as on understanding the viewer's psychological processes, for instance, by tracking eye movements (Boy et al. 2020) or analyzing the semantic and emotional content of YouTube comments (Amarasekara and Grant, 2019; Shapiro and Park, 2015). However, to the best of our knowledge, the use of affect in the communication of the scientific content itself has not been investigated as a potential driver of public engagement.

We seek to contribute to the literature by addressing two research questions: How is affect used in TED Talks in contrast to other science communication media, and is affect as a surface-level characteristic of science communication associated with audience engagement in the social media environment of YouTube? We adopt a data-driven approach to address these questions. Our analysis establishes, for the first time, affect as a potential driver of lay audience engagement with science communication on social media.

A database of TED Talk transcripts and engagement on YouTube

We downloaded all available transcripts and corresponding information (e.g., title, presenter, tags) of TED Talk videos from www.ted.com ($N = 6304$). In processing the transcripts, we eliminated any interview sections that followed the presentations. This also led us to remove 465 transcripts that consisted exclusively of interviews, leaving 5839 transcripts for further analysis. We obtained associated engagement data using the YouTube API and retrieved all available engagement data, which included the number of views, likes, dislikes, and comments, for all 3545 videos published on the TED YouTube channel. We then matched the transcripts and engagement data using the talk titles. Entries were matched using two strategies. First, we identified 2475 exact title matches. Then, we looked for matches in the remaining 1070 using approximate string matching and manual checking. This was necessary because many talks are published on YouTube using a different title than is used on the TED website. An additional 487 matches could thus be identified, amounting to a total of 2962 complete entries, all published between early 2007 and the end of 2020. The data were obtained on December 29th, 2020.

Identifying science in TED Talk transcripts. Although TED Talks are widely considered a form of science communication (e.g., Sugimoto and Thelwall, 2013), not all talks are science talks. While many TED speakers are academics, many others are, for instance, celebrities, journalists, athletes, and activists. In a study conducted by MacKrill et al. (2021) that examined TED Talks from 2006 to 2017, the authors found that only 27.4% of all talks were given by academics (i.e., people with a higher education degree and affiliated with a university). Past work on TED Talks has addressed the diversity of speakers and content by using the topic tags that TED assigned to each talk to characterize its content. For instance, Sugimoto and Thelwall (2013) used four of the 10 most frequent TED-assigned tags—"Science," "Technology," "Arts," and "Design"—to distinguish between the two topics Art & Design and Science & Technology.

Using a similar approach, we inferred topics from talk tags bottom-up using semantic network analysis (Kenett et al. 2020; Siew et al. 2019). Specifically, we used the co-occurrences of talk tags (e.g., "Physics" or "Medicine") to identify talk topics on the basis of homogeneous groups of tags (for a similar approach, see Wulff and Mata, 2022). In total, there were 447 tags; on average, 8.2 tags were assigned to each talk. Our approach to identifying science in TED Talks consisted of four steps. First, we determined the relatedness of each pair of tags using the Jaccard similarity. The Jaccard similarity measures the relatedness between tags by

relating the number of TED Talks for which the two tags co-occurred to the number of TED Talks for which either of the tags occurred:

$$J(A, B) = 1 - \frac{\left| A \bigcap B \right|}{\left| A \bigcup B \right|}.$$

Second, we used the relatedness of tag pairs to construct a weighted network of tags and apply the Louvain modularity detection algorithm as implemented in the igraph R package (Csardi and Nepusz, 2006) to identify homogenous groups of tags within the network (Blondel et al. 2008; Haslbeck and Wulff, 2020). Note that the Louvain algorithm compares favorably to other modularity and clustering algorithms (e.g., Emmons et al. 2016; Miasnikof et al. 2020; Pradana et al. 2020; Williams et al. 2019). The algorithm produced seven groups (hereinafter referred to as topics), which we labeled *Mind*, *Entertainment*, *Tech*, *Health*, *Cosmos*, *Environment*, and *Society*. Third, we substituted tags with their topic assignments and used the maximum positive point-wise mutual information, a common metric to assess the strength of semantic relationships (Bullinaria and Levy, 2007), between talks and topics to assign each talk to one of our seven topics.

To assess the quality of the mapping between talks and topics, we conducted a text analysis of talk titles. Using point-wise mutual information, we determined the most relevant words in TED Talk titles for each of the topics (see Fig. 1). The titles of talks assigned to *Mind* contained words such as "depressed," "compassion," and "decisions"; those assigned to *Entertainment* contained words such as "comedy," "poetry," and "violin"; those assigned to *Tech* contained words such as "hacked," "computers," and "net"; those assigned to *Health* contained words such as "synthetic," "diseases," and "antibiotics"; those assigned to *Cosmos* contained words such as "planets," "galaxies," or "Mars"; those assigned to *Environment* contained words such as "ocean," "trees," and "sustainable"; and those assigned to *Society* contained words such as "gun," "immigration," and "corruption." We further used a pre-trained sentence embedding, the Universal Sentence Encoder (Cer et al. 2018), to compare the semantic similarity of talk titles from the same topics to those of different topics and found that the within-topic similarity exceeded the between-topic similarity for every topic, with the difference in terms of Cohen's *d* ranging from 0.18 (*Entertainment*) to 0.69 (*Cosmos*) Together, these results indicate an accurate mapping of talks to semantically distinct topics.

Finally, to address the question of which set of TED Talks most concerns science communication, we computed a science index for each of the seven topics. This index reflected the percentage of talks in each of the seven topics that either were assigned the tag "Science" or contained the words "science," "experiment," or "study" in the transcript. Using this index, we found that the topic *Health* (79%) was most linked to science, followed by *Cosmos* (78%), *Mind* (69%), *Environment* (64%), *Tech* (58%), *Society* (43%), and *Entertainment* (37%).

Sentiment analysis. To address how affect is used in TED Talks compared with other science communication media, we relied on a dictionary-based approach (Denecke, 2008), a common approach to sentiment analysis (Feldman, 2013; Medhat et al. 2014). The approach involves mapping, wherever possible, the words in a text—in this case, the talk transcripts—to their corresponding sentiment value in the dictionary and calculating the sums of these values. In contrast to previous approaches, which often made use of the proprietary Linguistic Inquiry and Word Count (LIWC) database (e.g., Berger and Milkman, 2012; Brady et al. 2017; Hwong et al. 2017; Milkman and Berger, 2014), we relied on the openly available SentiWordNet sentiment dictionary (Baccianella et al. 2010). SentiWordNet contains more than 20,000 words with affect values ranging from −1 (most negative) to 1 (most positive). Like other sentiment dictionaries, SentiWordNet contains more negative (55%) than positive (45%) words, resulting in a negative average value of -0.06 ($SD = 0.34$). Using SentiWordNet, we calculated two sentiment summaries of the sentiment values *s* for each transcript. First, to capture whether the speaker used predominantly positive or negative words, we calculated an affective *valence* score,

$$valence = \frac{1}{n} \sum_i^n s,$$

where *n* is the total number of sentiment values available in a transcript. Second, to capture the speaker's tendency to rely on affect-laden words, irrespective of whether they have positive or negative valence, we calculated an affective *density* score,

$$density = \frac{1}{n} \sum_i^n I(s \neq 0),$$

where *I*() is an indicator function assigning a value of 1 when $s \neq 0$ and a value of 0 when $s = 0$ or unavailable. To our knowledge, the distinction between sentiment valence and density is a novel contribution, although related notions of sentiment density have been discussed in the literature (see Dong et al. 2013; Liu et al. 2018; Varshney and Wagh, 2017).

Dimensions of engagement. Past work seeking to quantify engagement on social media has mostly focused on combined engagement scores, calculated as a weighted sum of all available aspects of engagement (e.g., Hwong et al. 2017; Kim and Yang, 2017; Kujur and Singh, 2018; Vadivu and Neelamalar, 2015). Such approaches are sensible in light of typically strong correlations between engagement variables and can simplify matters in situations where the main goal is to generate a single metric capturing overall engagement. Recent investigations have, however, highlighted the value of distinguishing between different types of engagement. For instance, Srinivasan et al. (2013) found that image posts tend to garner more likes than comments, whereas the opposite was true for text posts. We decided against relying on a single engagement measure, to be able to detect relationships between affect and different forms of engagement. We therefore used a data-driven approach to extract independent

Fig. 1 Most relevant words in TED Talk titles per topic. Size of the words reflects the positive point-wise mutual information between the words and the topics derived from the network of tag co-occurrences. The 30 most relevant words per topic are displayed.
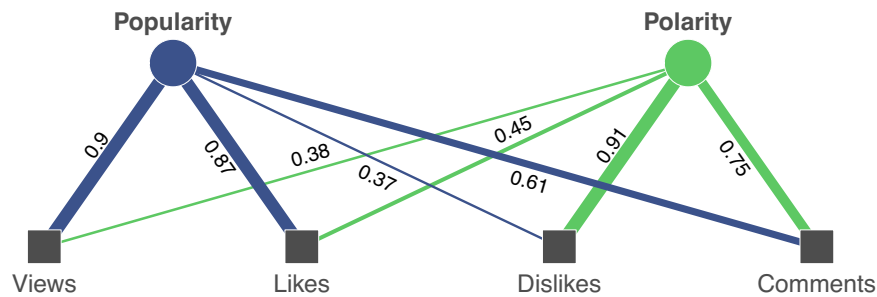
Fig. 2 Composition of engagement components. The figure shows the loadings of the four engagement variables on two principal components that constitute two forms of engagement: popularity and polarity. The presented solution accounts for 95.4% of the variance of the four manifest engagement variables.

engagement dimensions from the variables available. To do this, we applied principal component analysis to the four engagement variables accessible through the YouTube API: *views*, the number of times a video was clicked on; *likes*, the number of times viewers clicked the like button; *dislikes*, the number of times viewers clicked the dislike button; and *comments*, the number of times viewers left a comment. These variables were highly correlated ($0.70 < r < 0.92$), due to the fact that likes, dislikes, and comments are secondary to a video being viewed. Two engagement components were able to account for 95.4% of the total variance (see Fig. 2). The first engagement component, which we labeled *popularity*, captured positive reactions in the form of views and likes, whereas the second engagement component, labeled *polarity*, captured negative or contrarian reactions in the form of dislikes and comments.

Exploring the use of affect in TED Talks
Before analyzing how affect is linked to engagement, we took two approaches to gain insight into how it is used in TED Talks. First, we compared the values of affective valence and density in TED Talk transcripts to those in other text- or video-based media: a random subset of 1000 scientific articles on the preprint server arXiv[2], which primarily report research on STEM topics; a random subset of 1000 scientific articles from the journal *Psychological Science*[3], which report results on all topics in psychology, including research on emotion and affect; and random samples of text sources of other media, including Wikipedia articles, news articles, and subtitles of TV shows, soap operas, and movies[4]. This analysis revealed that the use of affect in TED Talks is distinct from all reference media (see Fig. 3A). TED Talks show considerably higher affective valence and, in particular, higher density than all text-based media (i.e., academic articles from arXiv and *Psychological Science* articles, books, Wikipedia articles, and news articles) but also lower affect and density than all video-based media (i.e., movies, TV shows, and soap operas). The analysis also revealed that the use of affect in TED Talks is, on average, more similar to that in other video-based media than that in text-based media, especially considering traditional expert-to-expert science communication in the form of academic articles. Nevertheless, there was also considerable variance in the use of affect in TED Talks, spanning the full gamut between the use of affect found in text and video-based media.

Second, we analyzed the valence and density of TED Talks as a function of the publishing year and topic in order to assess whether the use of affect in TED Talks has been stable over time and is independent of the topic (see Fig. 3B–E). This analysis revealed that the valence in TED Talks has decreased since 2007, whereas density seems to have increased, at least in recent years. Furthermore, the analysis showed that there were noticeable differences in the use of affect between topics. Affective valence

was most positive in *Entertainment*, followed by *Tech*, *Cosmos*, *Mind*, *Health*, *Society*, and finally *Environment*, whereas affective density was highest in *Mind*, followed by *Health*, *Tech*, *Environment*, *Entertainment*, *Society*, and finally *Cosmos*. We also analyzed the link between publishing year and topic and found that talks on *Society*, *Health*, and *Environment* have become more frequent at the expense of, in particular, talks on *Entertainment*, which may account for the temporal trends in the use of affect across time.

In sum, the language in TED Talks contains elevated levels of affect valence and density that are more similar to video-based than text-based media, including those reflecting expert-to-expert science communication in the form of academic articles. Furthermore, there was considerable variance in the use of affect in TED Talks, which is partially accounted for by differences in publishing years and topics.

The link between affect and engagement with TED Talks
To evaluate the role of affect in engagement with TED Talks, we ran separate regression analyses for our two engagement components (see Table 1). As predictors, we included valence and density as well as two sets of covariates: First, to control for the differences in the use of affect presented in the section *Exploring the Use of Affect in TED Talks*, we included the talk topic and date of publishing on YouTube. Including the publishing date also allowed us to control for differences in engagement, in particular concerning the number of views, which varies as a function of a video's age at data collection. Second, to control for factors other than affect that might drive engagement, we included the duration of the video and the Flesch Reading Ease score, which captures the accessibility of the language used in the talk (Flesch, 1948). The results are illustrated in Fig. 4. We found that more positive valence and higher density were associated with higher popularity. The effect of density ($d = 0.21$) was twice as large as the effect of valence ($d = 0.12$); however, both effects were small in magnitude. High popularity was also associated with long duration ($d = 0.32$), high readability ($d = 0.20$), and one topic, *Mind*. In contrast, *Environment* and *Society* were associated with low popularity. Polarity was associated negatively with valence ($d = -0.08$), but not density ($d = 0.02$). The effect of valence implies that more negative valences were associated with high polarity; however, this effect was small in magnitude. High polarity was also associated with longer duration ($d = 0.20$) and with the topic *Society*. *Health*, *Cosmos*, and *Environment*, in contrast, were negative in polarity.

To address whether the association of affect and engagement generalized across talk content, we compared the effects of affect on engagement for talks with a given topic or tag against talks without the topic or tag, using models that included all other predictors presented above except topic (see Table 1). In other words, we
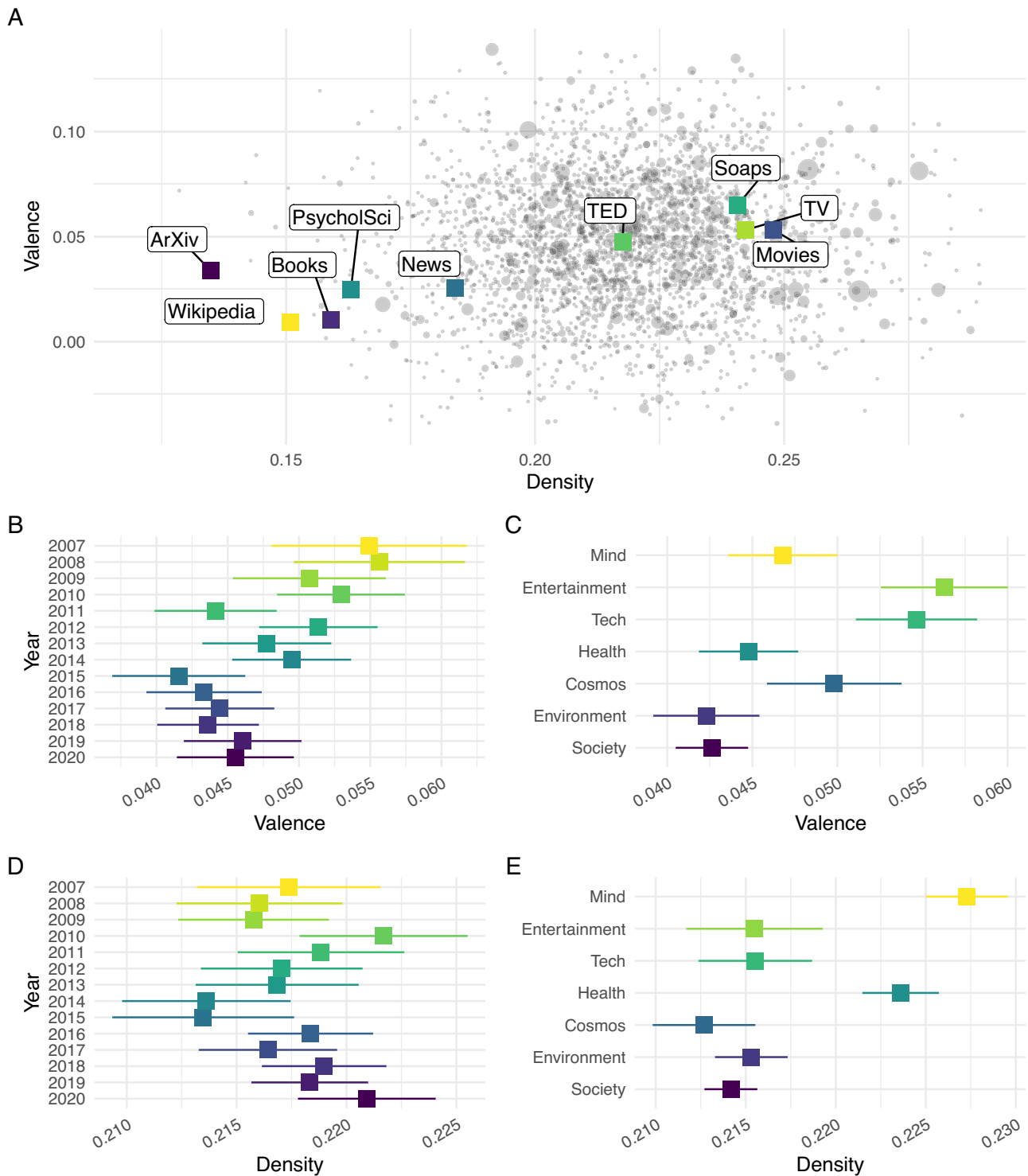
Fig. 3 Affect valence and density of TED Talks. A Circles in the background show the valence and density of all 2962 TED Talks, with their size scaled according to number of views. Squares in the foreground show the average valence and density of the TED Talks and the reference texts (see section *Exploring the Use of Affect in TED Talks*). **B–E** Average valence and density, shown separately for each publishing year and topic. For details on the topic extraction, see section *Identifying Science in TED Talk Transcripts*.

evaluated by how much and in which direction the content of talks moderated the effect of affect on engagement. Figure 5 illustrates this moderation in terms of Cohen's *d* for both engagement variables, popularity and polarity, and both content levels, topics and tags. There was considerable moderation for some but not all content. Beginning with popularity (Panels A and C), talks from

the topic *Environment*, especially those with the tags "Green" or "Sustainability," showed a noticeable reduction in the effect of density on popularity. The strength of the reduction implies that density in talks from the topic *Environment* was no longer related to popularity ($d = -0.02$). The opposite—an increase in the link between density and popularity—was the case for talks from the

topic *Mind*, especially those tagged with "Decision-Making" or "Mental Health." Furthermore, talks from the topic *Society* showed an elevated effect of valence, in particular those tagged with "Immigration" or "Refugees"; valence in these talks was more strongly related to popularity than was the case in talks from other topics. We observed the opposite for talks from the topic *Health*, in particular those tagged with "Medicine" or "DNA," with the result that valence was related mildly negatively to popularity ($d = -0.07$) within *Health*-related talks. Compared to these four topics, *Cosmos*, *Tech*, and *Entertainment* showed lower levels of moderation for popularity.

Turning to polarity (Panels B and D), talks from the topic *Tech*, especially those tagged with "AI" and "Machine Learning," and talks from *Environment*, especially those tagged with "Green" and "Sustainability," showed increased effects of density compared to talks from other topics, resulting in strong positive associations with polarity within these topics (*Tech*: $d = 0.49$, *Environment*: $d = 0.27$), whereas density in talks from *Society*, especially those tagged with "Refugees" and "Criminal Justice," showed a reduction in the effect of density on polarity, resulting in a small negative effect ($d = -0.17$). Furthermore, talks from *Entertainment* showed an increase in the effect of valence on polarity, with more positive valence being associated with a small increase in polarity ($d = 0.14$). In comparison, talks from the topics *Mind*, *Cosmos*, and *Health* showed smaller moderation effects for polarity.

Finally, we analyzed the association of affect and engagement with respect to the science index. We observed a small moderation effect for popularity, with talks of scientific content exhibiting a slightly reduced association of valence and popularity and a slightly increased association of density and popularity. For polarity, no moderation was observed. Consequently, the effect of

affect on engagement was largely unchanged for talks with a positive science index. Valence remained positively related to popularity ($d = 0.06$) and negatively related to polarity ($d = -0.07$), whereas density was more strongly related to popularity ($d = 0.29$) and unrelated to polarity ($d = 0.02$).

In sum, affective valence and density were significantly linked to engagement with TED Talks on YouTube: Increased valence and density were associated with increased popularity, and increased valence but not density were associated with negative polarity. These links were moderated by topic, with some topics seeing significantly pronounced or reversed relationships, suggesting that the link between affect and engagement depends in part on a talk's content. However, we did not observe meaningful moderation as a function of the science index, suggesting that the moderation by content is independent on whether the content focuses on science or not.

Discussion

Increasingly, scientists are communicating science to the general public. One example of this is TED Talks, where researchers give short presentations directed at a broad audience that are recorded and shared online. Effective science communication can thereby reach large audiences far beyond the scientific community. Here, we investigated the role of affect as a potential moderator of effective science communication in the context of social media, analyzing how affect expressed in the transcripts of TED Talks corresponds with engagement on YouTube. First, we observed that the use of affect in TED Talks in terms of valence and density is more similar to affect-laden visual media such as movies and soap operas than to traditional text-based media such as books, news articles, and academic articles. Second, we found that the two measures of affect were significantly related to two components of engagement: popularity and polarity. Higher affective valence was associated with higher popularity, reflecting more views and likes, and lower polarity, reflecting fewer dislikes and comments. Higher affective density, on the other hand, was related to higher popularity for almost all topics. Third, we observed substantial moderation of these effects by the topic of the talk, but not by whether the talk contained scientific or nonscientific content.

Our results demonstrate that affect as a surface-level characteristic of science communication on social media can impact how the public engages with scientific content. There are at least two potential explanations for this link. First, higher levels of affect may influence the affective state of the audience, e.g., heighten or lower its mood or arousal, and thereby impact

Table 1 Predicting the popularity and polarity of TED Talks.

	Popularity			Polarity		
	d	F	p	d	F	p
Affect						
Valence	0.12	9.68	0.002	−0.08	5.05	0.024
Density	0.21	32.21	<0.001	0.02	0.29	0.589
Control						
Duration	0.32	72.08	<0.001	0.20	28.36	<0.001
Readability	0.20	30.43	<0.001	−0.06	2.85	0.914
Topic	0.42	21.74	<0.001	0.43	22.58	<0.001
Date	−0.05	1.76	0.184	0.51	188.07	<0.001
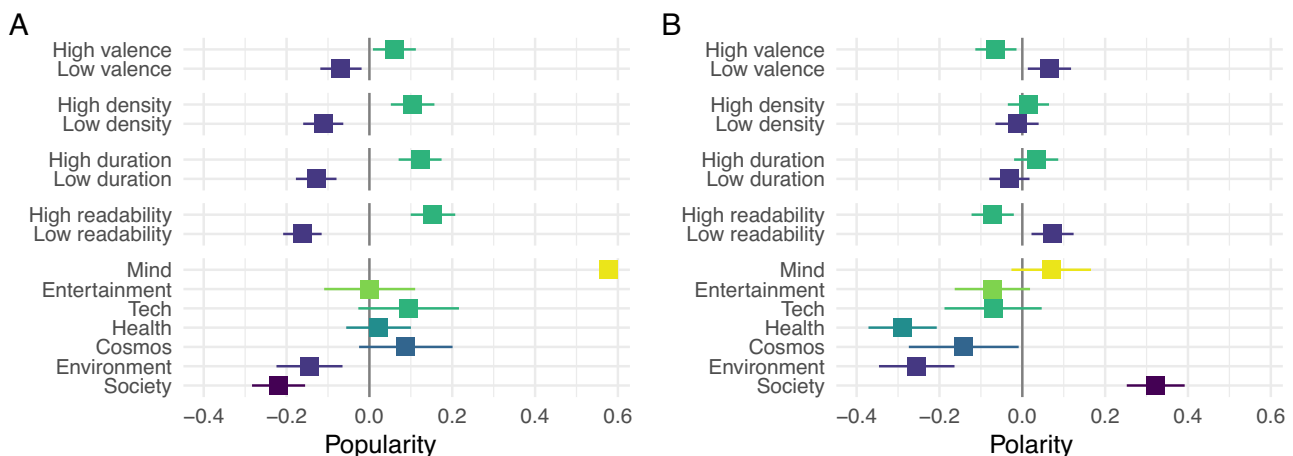
Fig. 4 Engagement as a function of the valence, density, duration, readability, and topic of TED Talks. The panels show the means of popularity (**A**) and polarity (**B**) for above- and below-median values in the predictors.
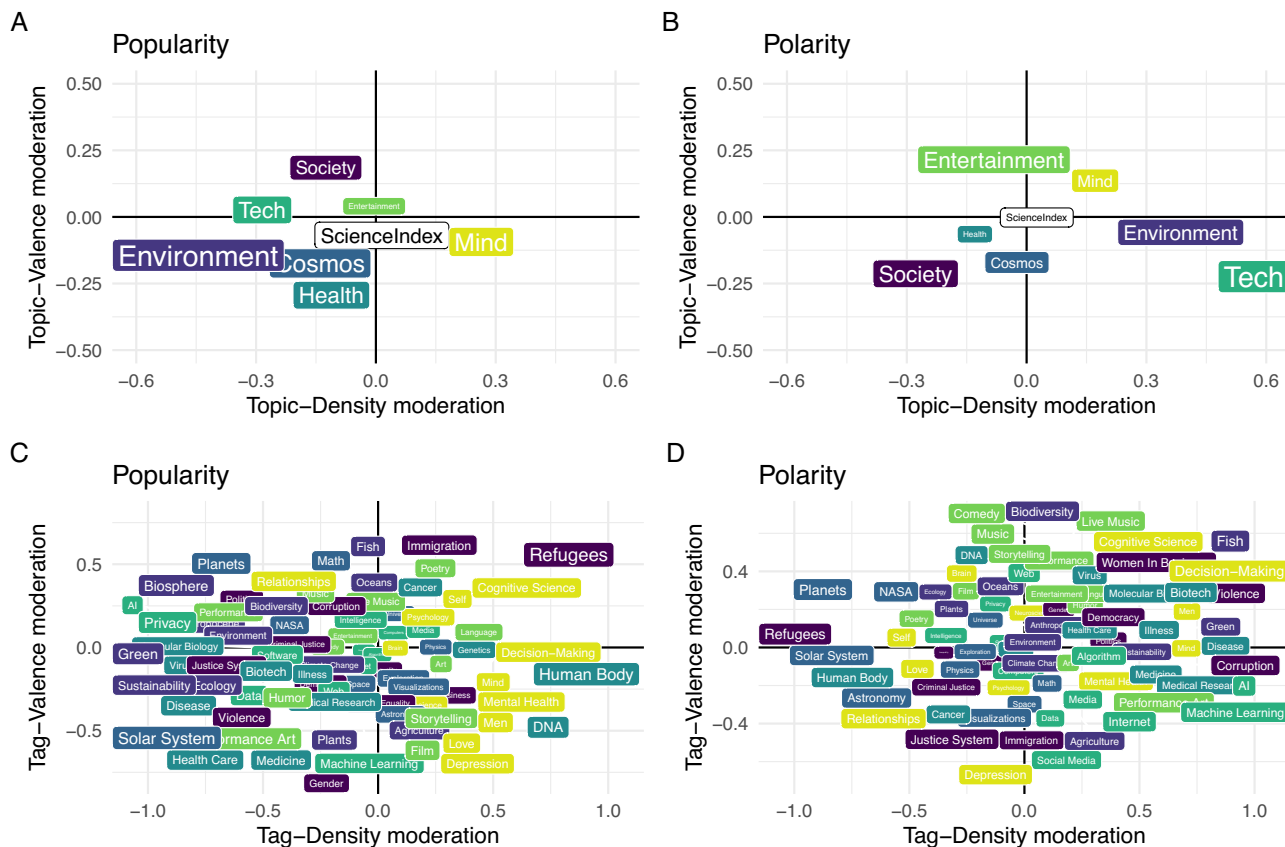
Fig. 5 Moderation of the effect of affect on engagement by talk topic and talk tag. The figure shows the difference in effect size *d* between the effects of affect on engagement for talks with given topic label or tag versus those without the topic or label. Positive values indicate that talks with the topic label or tag show a more positive effect than talks without the label or tag. Panels **A** and **C** show the results for models predicting popularity; Panels **B** and **D** show the results for polarity. Color indicates topic and size of the magnitude of moderation across both density and valence.

engagement. Second, higher levels of affect may signal more opinionated and assertive positions that increase the likelihood of engagement, whether supportive or critical. It seems plausible that both accounts are at least partially true. Associations of mood or arousal with engagement in social media are well documented (e.g., De Choudhury et al. 2012; Kujur and Singh, 2018; Osseweijer, 2006; Schreiner et al. 2021), and moderation of the association between affect and engagement was particularly pronounced for controversial or disruptive topics (e.g., "Refugees," "AI," "Sustainability," and "Health Care"), where the audience may favor an opinionated or a more measured approach (Hall et al. 2018; Hertwig and Wulff, 2021).

Our results may have practical implications for science communication on YouTube and similar social media outlets. They suggest that communicators can leverage the two components of affect to increase the public's engagement with their content on social media. Specifically, if science communicators incorporate more affect-laden words overall (i.e., higher density) and more positive rather than negative affect words (i.e., higher valence), their content may receive more views and likes (i.e., higher popularity) as well as fewer dislikes and comments (i.e., lower polarity as a result of higher valence). However, while increasing the valence and density of one's content may lead to an increase of its popularity on average, this effect may not generalize to all types of content. Making a talk more positive (i.e., higher valence) may backfire, for instance, when the talk already has high valence or when it does not meet the expectations of the audience. Furthermore, although a higher density of affect in talks is linked to higher popularity in almost all cases, simply increasing the

density of affect in a TED Talk without considering the overall use of language (e.g., jargon, visual imagery, story arc) may not yield the desired effects. Therefore, it is essential that science communicators understand that the way in which they communicate does indeed influence how their message is received and disseminated beyond the scientific community. In other words, to disseminate scientific findings to a broader audience, scientists may need—and are, perhaps, already expected—to become "fluent" in the many languages of science communication beyond traditional publications (e.g., blog posts or video essays; for a discussion of science communication in other formats, see Ho et al. 2021).

Our study has several limitations deserving of discussion. First, it relies on a purely correlative design. As a consequence, we can only speculate as to the causal mechanisms underlying our results and must refer to future experimental work to settle the issue. Second, TED Talks are but one form of public science communication (MacKrill et al. 2021; Sugimoto and Thelwall, 2013; Verjovsky and Jurberg, 2020). It is unclear to what extent our findings translate to, for instance, academic posts on social media (Rohrer et al. 2021) or traditional press releases, especially considering that text-based forms of science communication were found to rely less on affective language than TED Talks. Third, and relatedly, TED Talks are unusual in that they are used not only by academics to communicate scientific content, but also by other professionals to communicate ideas that may or may not relate to science. Scientific content in a nonscientific context may be evaluated differently than in a medium geared exclusively towards science communication. However, the presence of

nonscientific content is not unique to TED Talks and is likely to be found in most media used for public science communication. Fourth, the engagement variables available to us did not include shares—a stronger and more participatory form of engagement than the engagement variables in our analysis and a crucial aspect of content dissemination on social media (Shao, 2009). It is probable that shares would fall into our popularity component, given that they have been linked to higher ratings of scientific content's interestingness and usefulness—suggesting that shares often express support—as well as higher ratings of emotionality (a subjective measure similar to density) and positivity (an objective measure similar to valence; see Milkman and Berger, 2014). Accordingly, we expect that talks with higher valence and density would receive more shares.

All in all, our results establish an association between a TED Talk's affective content and engagement on social media along multiple dimensions of affect and engagement. Given the data-driven approach adopted in this investigation, we were unable to identify detailed mechanisms underlying the link between affect and engagement. However, it is possible, if not plausible, that affect codetermines engagement and reach among lay audiences on social media.

Data availability

The datasets generated during and analyzed in this study are available at this repository (https://osf.io/53cvg/).

Notes

1 Retrieved from socialblade.com on August 7, 2021.
2 Downloaded from www.kaggle.com/Cornell-University/arxiv
3 Downloaded from the journal's homepage.
4 Obtained from https://www.english-corpora.org/

References

Aldous KK, An J, Jansen BJ (2019) View, Like, Comment, Post: Analyzing User Engagement by Topic at 4 Levels across 5 Social Media Platforms for 53 News Organizations. Proc Int AAAI Conf Web Soc Media 13:47–57. https://doi.org/10.1609/icwsm.v13i01.3208

Allgaier J (2020) Science and medicine on YouTube. In Hunsinger, J., Allen, M. M. & Klastrup, L. (eds.) Second International Handbook of Internet Research, 7-27 (Springer Netherlands). https://doi.org/10.1007/978-94-024-1555-1_1

Amarasekara I, Grant WJ (2019) Exploring the YouTube science communication gender gap: A sentiment analysis. Pub Underst Sci 28:68–84. https://doi.org/10.1177/0963662518786654

Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Lrec 10:2200–2204

Berger J, Milkman KL (2012) What Makes Online Content Viral? J Mark Res 49:192–205. https://doi.org/10.1509/jmr.10.0353

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech 2008:P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Boczkowski P, Mitchelstein E, Matassi M (2017) Incidental news: how young people consume news on social media. In Proceedings of the 50th Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2017.217

Boy B, Bucher H-J, Christ K (2020) Audiovisual science communication on TV and YouTube. How recipients understand and evaluate science videos. Front Commun 5:608620. https://doi.org/10.3389/fcomm.2020.608620

Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ (2017) Emotion shapes the diffusion of moralized content in social networks. Proc Natl Acad Sci 114:7313–7318. https://doi.org/10.1073/pnas.1618923114

Brossard D (2013) New media landscapes and the science information consumer. Proc Natl Acad Sci 110:14096–14101. https://doi.org/10.1073/pnas.1212744110

Bullinaria JA, Levy JP (2007) Extracting semantic representations from word co-occurrence statistics: A computational study. Behav Res Methods 39:510–526. https://doi.org/10.3758/BF03193020

Cer D, et al. (2018) Universal sentence encoder. arXiv. https://doi.org/10.48550/arXiv.1803.11175

Collins K, Shiffman D, Rock J (2016) How are scientists using social media in the workplace? PLOS ONE 11:1–10. https://doi.org/10.1371/journal.pone.0162680

Covington P, Adams J, Sargin E (2016) Deep neural networks for youtube recommendations. In Proceedings of the 10th ACM Conference on Recommender Systems, 191–198. https://doi.org/10.1145/2959100.2959190

Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Systems 1695. https://igraph.org

Davidson J, et al. (2010) The YouTube video recommendation system. In Proceedings of the fourth ACM Conference on Recommender Systems, 293-296. https://doi.org/10.1145/1864708.1864770

Davies SR (2019) Science communication as emotion work: Negotiating curiosity and wonder at a science festival. Sci Culture 28:538–561. https://doi.org/10.1080/09505431.2019.1597035

Davies SR, Halpern M, Horst M, Kirby DS, Lewenstein B (2019) Science stories as culture: experience, identity, narrative and emotion in public communication of science. J Sci Commun 18:A01. https://doi.org/10.22323/2.18050201

De Choudhury M, Counts S, Gamon M (2012) Not all moods are created equal! Exploring human emotional states in social media. In Proceedings of the International AAAI Conference on Web and Social Media, 6, 66–73. https://doi.org/10.1609/icwsm.v6i1.14279

Denecke K (2008) Using SentiWordNet for multilingual sentiment analysis. In 2008 IEEE 24th International Conference on Data Engineering Workshop, IEEE, 507-512. https://doi.org/10.1109/ICDEW.2008.4498370

Dong R, Schaal M, O'Mahony MP, McCarthy K, Smyth B (2013) Mining features and sentiment from review experiences. In International Conference on Case-Based Reasoning, Springer, 59–73. https://doi.org/10.1007/978-3-642-39056-2_5

Emmons S, Kobourov S, Gallant M, Börner K (2016) Analysis of network clustering algorithms and cluster quality metrics at scale. PLOS ONE 11:1–18. https://doi.org/10.1371/journal.pone.0159161

Feldman R (2013) Techniques and applications for sentiment analysis. Commun ACM 56:82–89. https://doi.org/10.1145/2436256.2436274

Flesch R (1948) A new readability yardstick. J Appl Psychol 32:221–233. https://doi.org/10.1037/h0057532

Fronzetti Colladon A, D'Angelo CA, Gloor PA (2020) Predicting the future success of scientific publications through social network and semantic analysis. Scientometrics 124:357–377. https://doi.org/10.1007/s11192-020-03479-5

Gheorghiu AI, Callan MJ, Skylark WJ (2020) A Thin Slice of Science Communication: Are People's Evaluations of TED Talks Predicted by Superficial Impressions of the Speakers? Soc Psychol Personal. Sci 11:117–125. https://doi.org/10.1177/1948550618810896

Hall MG et al. (2018) Negative affect, message reactance and perceived risk: how do pictorial cigarette pack warnings change quit intentions? Tobacco Control 27:e136–e142. https://doi.org/10.1136/tobaccocontrol-2017-053972

Haslbeck JM, Wulff DU (2020) Estimating the number of clusters via a corrected clustering instability. Comput Stat 35:1879–1894. https://doi.org/10.1007/s00180-020-00981-5

Hertwig R, Wulff DU (2021) A description–experience framework of the psychology of risk. Perspect Psychol Sci. https://doi.org/10.1177/1745691611026896

Ho M-T, Ho M-T, Vuong Q-H (2021) Total SciComm: A Strategy for Communicating Open Science. Publications 9:31. https://doi.org/10.3390/publications9030031

Hoiles W, Aprem A, Krishnamurthy V (2017) Engagement and popularity dynamics of youtube videos and sensitivity to meta-data. IEEE Transac Knowl Data Eng 29:1426–1437. https://doi.org/10.1109/TKDE.2017.2682858

Hwong Y-L, Oliver C, Van Kranendonk M, Sammut C, Seroussi Y (2017) What makes you tick? The psychology of social media engagement in space science communication. Comput Hum Behav 68:480–492. https://doi.org/10.1016/j.chb.2016.11.068

Kaul L, Schrögel P, Humm C (2020) Environmental science communication for a young audience: A case study on the #EarthOvershootDay campaign on YouTube. Front Commun 5:601177. https://doi.org/10.3389/fcomm.2020.601177

Kenett YN, Beckage NM, Siew CS, Wulff DU (2020) Cognitive network science: A new frontier. Complexity 2020:e6870278. https://doi.org/10.1155/2020/6870278

Kim C, Yang S-U (2017) Like, comment, and share on facebook: How each behavior differs from the other. Pub Relat Rev 43:441–449. https://doi.org/10.1016/j.pubrev.2017.02.006

Kohler S, Dietrich TC (2021) Potentials and limitations of educational videos on YouTube for science communication. Front Commun 6:581302. https://doi.org/10.3389/fcomm.2021.581302

Kujur F, Singh S (2018) Emotions as predictor for consumer engagement in YouTube advertisement. J Adv Manag Res 15:184–197. https://doi.org/10.1108/JAMR-05-2017-0065

Liu Z, Zhang W, Cheng HN, Sun J, Liu S (2018) Investigating relationship between discourse behavioral patterns and academic achievements of students in SPOC discussion forum. Int J Distance Educ Technol 16:37–50. https://doi.org/10.4018/IJDET.2018040103

MacKrill K, Silvester C, Pennebaker JW, Petrie KJ (2021) What makes an idea worth spreading? language markers of popularity in ted talks by academics and other speakers. J Assoc Inform Sci Technol 72:1028–1038. https://doi.org/10.1002/asi.24471

Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: A survey. Ain Shams Eng J 5:1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Miasnikof P, Shestopaloff AY, Bonner AJ, Lawryshyn Y, Pardalos PM (2020) A density-based statistical analysis of graph clustering algorithm performance. J Complex Netw 8:cnaa012. https://doi.org/10.1093/comnet/cnaa012

Milkman KL, Berger J (2014) The science of sharing and the sharing of science. Proc Natl Acad Sci 11:13642–13649. https://doi.org/10.1073/pnas.1317511111

Osseweijer P (2006) A new model for science communication that takes ethical considerations into account: The three-e model: Entertainment, emotion and education. Sci Eng Eth 12:591. https://doi.org/10.1007/s11948-006-0058-z

Pradana C, Kusumawardani S, Permanasari A (2020) Comparison clustering performance based on moodle log mining. In IOP Conference Series: Materials Science and Engineering, vol. 722(1). IOP Publishing, 012012. https://doi.org/10.1088/1757-899X/722/1/012012

Rakedzon T, Segev E, Chapnik N, Yosef R, Baram-Tsabari A (2017) Automatic jargon identifier for scientists engaging with the public and science communication educators. PLOS ONE 12:1–13. https://doi.org/10.1371/journal.pone.0181742

Rohrer JM et al. (2021) Putting the self in self-correction: Findings from the loss-of-confidence project. Perspect Psychol Sci 16:1255–1269. https://doi.org/10.1177/1745691620964106

Schreiner M, Fischer T, Riedl R (2021) Impact of content characteristics and emotion on behavioral engagement in social media: literature review and research agenda. Electron Commer Res 21:329–345. https://doi.org/10.1007/s10660-019-09353-8

Shao G (2009) Understanding the appeal of user-generated media: a uses and gratification perspective. Int Res 19:7–25. https://doi.org/10.1108/10662240910927795

Shapiro MA, Park HW (2015) More than entertainment: Youtube and public responses to the science of global warming and climate change. Soc Sci Inform 54:115–145. https://doi.org/10.1177/0539018414554730

Sharon AJ, Baram-Tsabari A (2014) Measuring mumbo jumbo: A preliminary quantification of the use of jargon in science communication. Pub Underst Sci 23:528–546. https://doi.org/10.1177/0963662512469916

Siew CS, Wulff DU, Beckage NM, Kenett YN (2019) Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. Complexity 2019:e2108423. https://doi.org/10.1155/2019/2108423

Srinivasan BV et al. (2013) Will your facebook post be engaging?Proceedings of the 1st Workshop on User Engagement Optimization - UEO '13, ACM Press, 25–28. http://dl.acm.org/citation.cfm?doid=2512875.2512881

Sugimoto CR, Thelwall M (2013) Scholars on soap boxes: Science communication and dissemination in TED videos. J Am Soc Inform Sci Technol 64:663–674. https://doi.org/10.1002/asi.22764

Vadivu VM, Neelamalar M (2015) Digital brand management - A study on the factors affecting customers' engagement in Facebook pages. In 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), IEEE, Avadi, Chennai, India, 71-75. http://ieeexplore.ieee.org/document/7225392/

Varshney V, Wagh RS (2017) Weighted sentiment score formulation using sentence level sentiment density for opinion analysis. Int J Comput Intell Res 13:285–298

Verjovsky M, Jurberg C (2020) Spreading ideas: TED talks' role in cancer communication and public engagement. J Cancer Educ 35:1206–1218. https://doi.org/10.1007/s13187-019-01583-6

Williams N et al. (2019) Comparison of methods to identify modules in noisy or incomplete brain networks. Brain Connect 9:128–143. https://doi.org/10.1089/brain.2018.0603

Wulff DU, Mata R (2022) On the semantic representation of risk. Sci Adv 8:eabm1883. https://doi.org/10.1126/sciadv.abm1883

Acknowledgements

Author contributions
OF, LTJ, and DW designed the study. DUW analyzed the data. OF, LTJ, and DW wrote the first draft of the manuscript and revised the manuscript.

Funding

Competing interests
The authors declare no competing interests.

Ethical approval
This article does not contain any studies with human participants performed by any author.

Informed consent
This article does not contain any studies with human participants performed by any author.

Additional information
Correspondence and requests for materials should be addressed to Dirk U. Wulff.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.