BERJ www.berjournal.com ISSN: 2619-9491 BUSINESS AND ECONOMICS RESEARCH JOURNAL

Business and Economics Research Journal Vol. 14, No.4, 2023 pp. 445-464 doi: 10.20409/berj.2023.425

Drivers of Popularity of Online Information: Content, Context and Psychological Processes

M. Utku Özmen[1], **M. Eray Yücel**[2]

Abstract: *This paper aims to study how people utilize (search for, choose, process, and evaluate) information provided on online domains, emphasizing the balance between context identifiers and the actual content of information and the psychological processes. The study assesses the popularity of online provided materials, TED Talks, in relation to the length of information, user ratings, and several content-related features. The paper employs a comprehensive naturalistic data set that covers the titles, duration, viewer-assigned ratings/tags, transcripts, various content identifiers, and popularity (number of views) of 2685 TED Talks. The results reveal the relevance of both content and context-related factors, as well as psychological processes, on the popularity of the talks. On the context side, using certain words in the title and the text, optimizing the talk pace and the length of the talk; on the content side, carefully incorporating rhetorical features are major factors that influence the popularity of the talks. On the psychological processes front, the popularity of talks is associated with positive emotions and anxiety among affective processes, and insight and tentativeness among cognitive processes.*

Keywords: TED Talks, Online viewing behavior, Popularity, Content and context, Psychological processes, LIWC, Economics of online information

JEL: D91, L86

Received : 06 January 2023
Revised : 30 March 2023
Accepted : 15 September 2023

Type : Research

1. Introduction

The widespread diffusion of the Internet and digital technologies have substantially transformed how individuals search for and process information, opening the way to the economics of attention and introducing essential challenges for the economics of online information provision. From a historical perspective, a student of sciences of the Late Medieval Era would be stunned and possibly confused if she/he was hypothetically beamed to 2023. Almost the same would apply to another scholar from the late 19th century, both amazed by the tremendous volume of information of several sorts available even through a hand-held communication device. While the visitor from Late Medieval times would be overwhelmed mainly by the vast amount of information spread over almost every single town, the visitor from the late 19th century would admire the state-of-the-art cross-referencing facilities and search algorithms installed everywhere. So, it is fair to suppose that they would suffer from “a poverty of attention” once they face “a wealth of information,” as suggested by Herbert A. Simon (1971). Would they complain about the nature of intellectuality, which possibly changed into “fast-moving consumption, or would they be critical of the observation that context often beats content?” In this paper, we aim to communicate our findings associated

Cite this article as: Özmen, M. U., & Yücel, M. E. (2023). Drivers of popularity of online information: Content, context and psychological processes. *Business and Economics Research Journal, 14*(4), 445-464. http://dx.doi.org/10.20409/berj.2023.425

[1] Lecturer, PhD., TOBB University of Economics and Technology, Faculty of Economics and Administrative Sciences, Department of Economics, Ankara, Turkiye, utkuozmen@gmail.com (Corresponding Author)

[2] Lecturer, PhD., Ihsan Dogramaci Bilkent University, Faculty of Economics, Administrative, and Social Sciences, Department of Economics, Ankara, Turkiye, eray@bilkent.edu.tr

with these questions in the context of the popularity of online information and in an environment where digital technologies have transformed user behaviour, considering the psychological processes that may be at play.

Özmen and Yücel (2019) state that a structural change in the accessibility and abundance of information has challenged the classical methods of knowing things at the nexus of content and context of information. This is based on the supposition that “the making of the intellectuality itself relied on finding an ambitiously fine line to separate quality from quantity, challenging from digestible, scientific from bogus and creative from straightforward”. The undeniable and permanent change of the parameters in the economics of attention, indeed, calls for several dimensions like the practically non-existing pecuniary marginal costs of information acquisition, so increased democratization of access to information, fast reproduction, and re-dissemination of information, so a radically evolved mannerism and eventual code of communication. Online retrieval of information is moneywise free (at the point of actual use by its audience, despite the existence of several paywalls for multi-tiered or varying quality information disseminated in certain online domains, which are not directly relevant here), leaving the amount of attention as the price of knowing things. Is this price so high when faced with a generous spectrum and volume of information? Is it too high to pay to locate the intellectual fine line between context and its inherent content? How do psychological processes mediate the selection of what to attend given the tremendous flow of information, and how much to appreciate that information? Putting the question forth provocatively: should we fear losing the traditionally established ways of intellectualism and how psychological processes affect that?

We resort to the “popularity of a specific piece of information” as our central measure to seek answers to intertwined questions posed. As a proxy rather than a direct measurement, popularity functions well as a reflection of applause or appreciation once we failed to pinpoint a better measure in our preliminary investigations. Then, we research the answers to the questions mentioned above by relating popularity to its potential covariates, i.e., we try to see how the context- or content-related factors, along with psychological processes, help explain the popularity. Focusing on popularity also sheds light on the eonomics of online information provision, as the providers of more popular content are more likely to enjoy the economic returns of their visibility and outreach.

We use the TED Talks as the setting of our analysis. TED -abbreviated from Technology Entertainment and Design- is a non-profit organization that has been organizing conferences on technology, science, and design-related topics since 1984. Currently, the ideas disseminated as short talks -labeled TED Talks- cover a more comprehensive range of topics, from science and technology to emerging global issues. The talks have been uploaded since 2006 on the website http://www.ted.com for free. In the empirical analysis, we use the titles, duration, and viewer-assigned ratings of more than 2,000 TED Talks to reach our statistical conclusions in the first place. Upon these, we further (1) consider other mechanical measurements about talks, like the word count and the pace of talks; (2) delve deeper into the transcripts of the talks by generating indicators of the actual content, like the use of specific phrases in and the psychological processes dominant in the talks and (3) integrate the view counts of the talks as a proxy for their popularity. All these variables are obtained directly from the database of TED Talks, while the content analysis is made by the Linguistic Inquiry and Word Count (LIWC) text analyzing tool.

While Özmen and Yücel (2019) quantify (1) the relations between viewer-assigned ratings, talk durations, and wording of talk titles and (2) the relations between previously assigned audience ratings and the subsequent online viewing behaviour, this study goes beyond that by unveiling the impacts of actual content, including the psychological processes, as well as the popularity of the talk. So, this paper tells more on the questioned balance between context and content in determining the “importance” of an information stream. Thus, this study's novelties are the coverage of the directly content-related measures and denoting popularity as the primary variable of interest. From the theoretical background, the paper builds on the information retrieval and processing theories and economics of attention as thoroughly discussed in Özmen (2015) and Özmen and Yücel (2019); and uses the online information provided in various formats -videos, and text- as the metrics for the empirical analysis.

The paper has practical implications for both online content providers seeking to maximize their outreach and for practitioners of big data analytics. On the context side, using certain words in the title and the text, optimizing the talk pace and the length of the talk; on the content side, carefully incorporating rhetorical features and emphasizing specific psychological processes are major factors that influence the popularity of the talks. Regarding the empirical methodology, the paper motivates how multivariate regression analysis is a more helpful approach than simple correlation analysis, a common practice in recent platforms analyzing big data, such as Kaggle, when using internet data for empirical analysis.

The remainder of the study is organized as follows: the next section gives the background of current research. Section 3 presents our empirical setup and results. Section 4 is dedicated to discussing findings before concluding our work in Section 5.

2. Background

Our central pillar to grasp the information-processing behavior of humans stems from cognitive psychology. This field provides us with several explanations of attention, like selective concentration. Facing many stimuli, the individual filters the unwanted ones to minimize her subsequent cognitive effort. Broadbent's (1958) bottleneck theory stating that excessive information that cannot be handled by one is ignored, and Treisman's (1964) modification of Broadbent stating that at the early stage, the available set of stimuli is processed in a parallel manner, while the selection occurs at a later stage, lay down the basics well. Late selection is also studied by Deutsch and Deutsch (1963) and Norman (1968).

In line with these key studies, the pertinence of information induces filtering and selection to occur later, calling for an active processing strategy defined by a person's goals. In an information-abundant environment, as experienced today, selective attention refers to attending to information that maximizes utility with respect to some objectives. Miller et al.'s (1960) information processing theory defines the "test-operate-test-exit" as the primary behavior unit. Information is processed sequentially, where an input starting the process is tested based on internal criteria, operated, and then tested again until a designated goal is reached.

According to Simon (1971), the availability of too much information results in poverty of attention, suggesting a need for efficient allocation of attention. Kahneman (1973) points to a possible upper limit for resources, including a person's attention to a task. Type of information, psychological state, enduring dispositions, and monetary intentions might be related to this upper limit. Lanham (2006) subsequently re-established the foundations by eloquently pointing at the fact that relative scarcities of information itself and the psychological effort to attain it, namely attention, are switched once easy reach of digital information has been granted. So, the economic aspects of the topic are to be understood in a different light welcoming the role of information providers as advertisers. Then, advertising increases the quality perceived by the users through initial stimuli, as discussed by Huberman (2009) and Simola et al. (2014). In direct relationship to this paper, earlier research efforts on online viewing behavior and TED Talks are not poor, as summarized in Table 1.

Table 1. An Overview of the Recent Related Literature

A. Selected Studies on Online Viewer Behavior	
Study	*Topics covered*
Segev and Ahituv (2010)	Dependence of online information viewing attitudes on culture and the organization of societies; differentiation of socio-political and entertainment concerns
Fiksdal et al. (2014)	Information saturation and fatigue as main reasons for stopping information retrieval
Wook and Salim (2014)	Specification requirements for visual aspects of information provision as the use of space, organization of information, and function and use of color
Lee et al. (2015).	Viability of predictive models of web browsing behavior based on its records

Table 1. An Overview of the Recent Related Literature (Continue)

B. Selected Studies on Online Viewer Behavior	
Study	*Topics covered*
Gao and Bai (2014)	Importance of the information provision style and consumers' perception
Khan and Vong (2014)	Effect of contextual features of the videos and the indicators related to the outreach of the users uploading on YouTube on the virality of videos
Özmen (2015)	Attractiveness of photos, tags, the dominance of visual content over the textual, and the attention-augmenting role of photos
Budzinski and Gaenssle (2018)	Content uploading behavior of social media superstars
Tafesse (2020)	Investigation of the link between the contextual features of trending YouTube videos with the number of viewings
Liao et al. (2021)	Recommendation systems, analysis through clustering analysis
Skarpa and Garoufallou (2021)	Public's information-seeking behavior on COVID-19 via surveys
Azzopardi (2021)	Cognitive biases in the field of information retrieval
Gordon et al. (2022)	Utilization of resources and strategies to seek, find, and use scholarly information and news, analyzing through a sample of physicists
Reisoğlu et al. (2022)	Impact of procedural and metacognitive processes on online search behavior
C. Selected Studies Focusing on TED Talks	
Study	*Topics covered*
Lopes et al. (2011)	Acoustic and speech recognition aspects
Cettolo et al. (2012)	Formation of a Web inventory of transcribed and translated talks
Rubenstein (2012)	Innovative future possibilities for TED Talks concerning mainly education
Pappas and Popescu-Belis (2013)	Utility for a one-class collaborative filtering task such as bookmarking
Sugimoto et al. (2013)	Linkages between academic citations and TED appearances
Di Carlo (2014)	Assessment of TED Talks via Hyland (2010)'s concept of proximity
Romanelli et al. (2014)	Benefits and and downsides of TED Talks for dissemination of ideas
Rousseau et al. (2014)	Use of TED data in language modelling
Tsou et al. (2014)	Impacts of presenter characteristics and platform on the reception of videos
Bertero and Fung (2017)	Algorithms to detect emotions in TED Talks
Chen and Lee (2017	Prediction of the audience's laughing behavior at TED Talks
Tanveer et al. (2018)	Relation between narrative trajectories and the ratings of TED Talks
Özmen and Yücel (2019)	Relationships between the talk duration and viewer ratings, attention driving factors; between the ex-ante wording of talk titles and ex-post viewer ratings
Kinnaird and Laudun (2019)	Discussion of how the talk transcripts are useful in examining accuracy and quality
Schwemmer and Jungkunz (2019)	Analysis of TED talk transcripts to investigate whether female speakers, speakers from different ethnic groups, and some topics are less represented on the stage
Gheorghiu et al. (2020)	The link between the scientific quality and entertainment value evaluations of the TED Talks with the speaker's characteristics
Pozdena (2020)	Cultural characteristics, content of the talk, and the appreciation of the audience
Almaged (2021)	Linguistic mechanism of disseminating knowledge the interface between knowledge, meaning, and social practices in terms of text and context
MacKrill et al. (2021)	Impacts of the language used in TED Talks on the popularity and viewer ratings
Fischer et al. (2021)	Affect valence, density, and polarization in TED talks
Liou and Tseng (2022)	Linguistic features and vocabulary in TED Talks
Wingrove (2022)	Coverage of academic lexis in TED Talks in comparison to that in academic lectures

Regarding the popularity of visual online content, previous studies mainly focused on YouTube. For instance, Khan and Vong (2014) analyze the virality of the videos. They explore the effect of contextual features of the videos, such as duration and category, and the indicators related to the characteristics of the users uploading videos, such as age, gender, and the number of subscribers on the virality of videos. They find that links, hit counts, and fan base are positively related to the number of viewings. Tafesse (2020) investigates how the contextual features of trending YouTube videos, such as title, tags, and descriptions, affect the number of viewings. One interesting finding of Tafesse (2020) is that titles with negative emotional sentiments attract more viewers.

On the other hand, the need for popularity is also an essential driver for online content providers. Lim et al. (2015) find that micro-bloggers with a low sense and expression of real themselves have a higher need for popularity. Budzinski and Gaenssle (2018) analyze the "social media superstar" phenomena, where experience in the market and upload frequency of content help explain the presence of social media superstars. The popularity of information providers may also further attract consumers. For instance, Mou and Shin (2018) find that online health advice-seekers associate the social popularity of the practitioner with trust and perceived quality.

A few studies engage with the popularity of TED Talks. Liu et al. (2017) study how rhetorical aspects of the talk generates a higher appreciation of the audience captured by the applause received. They find that sentences with more logical expressions and less personal pronouns, talks focusing more on the present than the past, and speakers who adopt a less formal tone are more likely to generate applause. Tur et al. (2018) study how the speaker's charisma affects the number of views and the ratings of the TED Talks. They find that more use of charismatic signals such as metaphors and contrasts increases the popularity of the talks (number of viewers) and the probability of the talk being rated inspiring. Maeno and Maeshiro (2018) try to predict the occurrence of a standing ovation at the end of the TED Talks by applying machine learning methods to the textual content of the talks. Meanwhile, MacKrill et al. (2021) investigate the link between the popularity of TED Talks and the use of language using the LIWC software, but they only focus on a limited number of indicators from the LIWC and only rely on simple correlation without controlling for potential confounding factors.

Focusing on the order of the events, for the empirical analysis, note that the first set of viewers is not informed about the content-related factors but they can only consider the context-related factors of the talk (such as title, ratings, duration, venue) as they are already available before viewing the talk, to decide on which talks to attend. However, later viewers can benefit from the sharings, postings and comments of the earlier viewers about the talk -word of mouth- which is an important aspect of the popularity. So, in that perspective, while there is a more direct relationship between context-related factors and popularity, the link between content-related factors and popularity occurs indirectly. The mechanism investigated in this analysis is that the context-related factors first attract the viewers. Later on, the interaction of the viewers with others through posts and comments about the characteristics of the talk such as the topic, presenter's style, and pace, further attract new viewers. Overall, controlling for relevant factors, including those related to content, and using a set of fixed effects enables a better estimation of the relationship between the context-related factors and popularity. Meanwhile, the indirect relation between content and popularity enables an ex-post evaluation of the association between content-related factors and popularity.

Given our research question and the earlier literature on the economics of attention and cognitive psychology, the Internet-based collection of TED Talks provides a suitable empirical environment. Even though it lacks the specific characteristics of a full-fledged experimental setup, the empirical strategy we elaborate in the next section allows us to extract a reliable body of information from contextual data and the transcripts of the talks to analyze their popularity.

3. Data and Empirical Analysis

This section is devoted to the presentation of our empirical analysis. In that, we first present the features of our data set. Then we explain how we obtained the content-related measurements. Upon these, we describe the estimation strategy and equation employed, immediately followed by its basic estimates. Rhetorical and psychological elements are further incorporated toward the end. Owing to its blending of several ingredients, the empirical analysis below might better serve the reader through an old-school linear reading.

3.1. Nature of the Dataset

Due to the impracticality of conducting a controlled experiment in the current context, we considered the talks enlisted at TED Talks website. So, from a statistical standpoint, our data set is naturalistic as we use data from all users in a natural setting. The dataset covers all 2,685 online talks from June 2006 (earliest) to February-2018. Most of the talks are in English, and all the talks have English transcripts. Note that the data covers only talks registered at the Ted Talks, and excludes TEDex talks which are locally and separately organized events in local languages. The data are retrieved from the public webpage, where all the talks are listed in subsequent sub-pages. Thus, we collect the following context-related information from these public pages: the presenter's name, title of the talk, date of the talk (month/year), duration of the talk (in minutes), and the top-two ratings associated with each talk. The ratings are adjectives rewarded by the viewers of the talks and are chosen from the available list, including courageous, inspiring, jaw-dropping, informative, beautiful, persuasive, fascinating, confusing, obnoxious, ingenious, ok, longwinded, unconvincing, and funny. For the data period, the top-two ratings awarded to each talk were also reported on the main page of the TED Talks. Only eight of these adjectives were given as top-two ratings for the talks in the sample, as seen in Table 2. The way the data is organized is a nice example of attribute-based information provision, where the title, duration, and ratings constitute the attributes.

As the first novelty of the paper, we incorporate the total (the log of total number of views: LVIEW) and per-unit-time (here, the log of views per month: LVPM) counts of viewing for each talk in our dataset in addition to those listed above. Compared to Özmen and Yücel (2019), where talk durations and ratings were linked to their determinants via Least Squares regressions, these counts considerably enrich our grasp of the data as they constitute our dependent variables (that quantify popularity) subsequently.

The second novelty of this paper stems from the comprehensive handling of content-related elements. Having web-scraped full transcripts of the TED Talks included in our data set, we are first able to employ quantitative indicators of the overall tone, psychological processes, and several rhetorical features of talks. Secondly, we can include in our analysis some features like the pace of the speaker as measured by "words per minute", the complexity of a talk as measured by "words per sentence" or "number of words longer than six letters", the emotional gestures triggered by a talk as measured by the occurrences of applause and laughter. The content-related information is generated by executing the Linguistic Inquiry and Word Count (LIWC) program -designed to analyze the usage of words in texts- on the transcripts of the talks (the details of the LIWC are provided in the next section). Such availability of content-related measurement elements allows us to study the balance between context and content and the role of psychological factors. The descriptive statistics of the main variables are presented inTable 2. Meanwhile, the complete list and the descriptive statistics of all LIWC indicators are presented in the Appendix, Table A1.

Table 2. Descriptive Statistics of Main Variables in the Analysis

Variable	Label	Obs.	Mean	Std. Dev.	Min	Max
Popularity:						
ln (View per month)	*lvpm*	2,130	10.25	1.19	7.77	13.04
ln (Total views)	*lview*	2,130	14.03	0.71	11.67	17.72
Main Ratings of the Talk:						
Rating_Beautiful	*Beautiful*	2,130	0.14	0.35	0	1
Rating_Courageous	*Courageous*	2,130	0.12	0.33	0	1
Rating_Fascinating	*Fascinating*	2,130	0.31	0.46	0	1
Rating_Funny	*Funny*	2,130	0.08	0.27	0	1
Rating_Informative	*Informative*	2,130	0.51	0.50	0	1
Rating_Ingenious	*Ingenious*	2,130	0.11	0.31	0	1
Rating_Inspiring	*Inspiring*	2,130	0.55	0.50	0	1
Rating_Persuasive	*Persuasive*	2,130	0.19	0.39	0	1
Talk duration and pace:						
Duration of talk (min)	*time*	2,130	13.71	3.83	6.02	20.08
Words per minute	*wpm*	2,130	150.97	32.93	0.31	248.72
Words per sentence	*wps*	2,130	16.72	4.30	6.00	97.67

Table 2. Descriptive Statistics of Main Variables in the Analysis (Continue)

Variable	Label	Obs.	Mean	Std. Dev.	Min	Max
Main LIWC Dimensions:						
Analytical Thinking	*analytic*	2,130	63.20	14.89	14.12	99.00
Clout	*clout*	2,130	80.40	9.80	29.84	99.00
Authentic	*authentic*	2,130	28.77	13.23	1	83.66
Emotional Tone	*tone*	2,130	52.75	20.15	1	99.00
Affect Words	*affect*	2,130	3.99	1.46	0	12.61
Social Words	*social*	2,130	10.44	3.84	0	33.33
Cognitive Processes	*cogproc*	2,130	11.21	2.54	0	20.06
Perpetual Processes	*percept*	2,130	2.88	2.15	0.19	25.00
Biological Processes	*bio*	2,130	1.74	1.38	0	9.08
Core Drives and Needs	*drives*	2,130	8.09	2.30	0	16.67
Relativity	*relativ*	2,130	13.28	2.31	0	21.37
Informal Speech	*informal*	2,130	0.42	0.40	0	6.12

Notes: The analysis considers talks with typical lengths -6- to 20-minute-long-. Therefore, we use observatiaons from 2130 talks out of 2685 in the estimation samples.

3.2. Specifics of Content-Related Measurements

It would be helpful to present the qualities of the specific software we benefited from, owing to its importance in quantifying the content-related features of talks. We used Linguistic Inquiry and Word Count (LIWC) software's 2015 edition in data preparation. It is a research-based software with commercial availability, the revenues of which are transferred/donated to the Department of Psychology of the University of Texas at Austin (Pennebaker, Boyd, Jordan and Blackburn, 2015; Pennebaker, Booth, Boyd and Francis, 2015). Using built-in dictionaries, LIWC evaluates a text by comparing the words in that text with the list of words in a specific dictionary and yields several summary indicators regarding the use of language in the text under the following headings (Pennebaker, Boyd, Jordan, and Blackburn, 2015):

- Word count
- Summary language variables
- Linguistic dimensions
- Other grammar
- Psychological processes

3.3. Estimation Strategy

Our estimating equation takes the form of Equation 1. Here y_i is the dependent variable (either $LVPM$ or $LVIEW$), $x_{li}, l = 1,2,\dots,L$, are the key explanatory variables (Beautiful, Courageous, Fascinating, Funny, Informative, Ingenious, Inspiring, and Persuasive, as in Özmen and Yücel (2019), and $z_{ki}, k = 1,2,\dots,K$, are other comprehensive explanatory variables containing various content-based indicators of rhetorical features and meaning. In addition to these, $D_{mi}, m = 1,2,\dots,M$, are the dummy variables for the year, month, and location of the talks. These combine into Equation 1, where $\epsilon_i, i = 1,2,\dots,2130$, are the Gaussian statistical error terms:

$$y_i = \alpha_0 + \sum_{l=2}^{L} \alpha_l x_{li} + \sum_{k=1}^{K} \beta_k z_{ki} + \sum_{m=1}^{M} \gamma_m D_{mi} + \epsilon_i \qquad (1)$$

The specificaiton 1 is estimated using the Least Squares (LS, OLS) with robust coefficient covariance matrices. Considering the x variables, x_1, our base (or reference) rating category -Beautiful- has been excluded from the specification.

In the following subsections, we report our estimates with their implications. In each, we first provide an account of the results based on a model where $LVPM$ is the dependent variable. Then, we extend our discussion to cover the cases where $LVIEW$ replaces $LVPM$. The main departure of our estimation strategy from that of Özmen and Yücel (2019) is the treatment of popularity as the dependent variable, rather than talk duration or rating. So, talk durations and ratings are used as right-hand side variables.

3.4. Basic Estimates

Our first set of estimates (called here the basic estimates) combines the popularity variables (LVPM and LVIEW) with the viewer rating variables of Özmen and Yücel (2019), talk duration (time), word per minute (wpm), and word per sentence (wps) measures for each talk. The basic estimates aim to assess the explanatory power of viewer-assigned ratings on the popularity of talks in various settings and to establish a base for our subsequent specifications. The results of the basic estimates are provided in Table 3.

In our elaborations, we maintain a simple notation to refer to empirical specifications, where (T-S) shows the S-th column of T-th table. Viewer rating variables are given in double-quotes, and we note the sign of estimated coefficients in parentheses. Using this notation, for instance, in (3-1), talks rated "courageous" (+) and "inspiring" (+) have statistically higher count of views, meanwhile talks rated "ingenious" (-) and "persuasive" (-) have statistically lower count of views compared to the base category of "beautiful". Talks rated "fascinating", "funny" and "informative", on the other hand, do not receive more count of views compared to talks rated "beautiful". We note that only for the rating categories, the comparison is always against a base category, which is "beautiful" in our case. That is, statistically (+) and (-) coefficients for ratings point to statistically higher or lower count of views for those talks rated as such compared to talks rated as beautiful. For all the other regressors, the interpretation is directly related to the count of views. After including talk durations (time) in (3-2), the rating categories maintain their signs and statistical significance. Talk duration has a significant negative impact on LVPM. In (3-3) the pattern in (3-2) is preserved with the addition of "informative" (+) as a significant regressor. Here, wpm has an insignificant negative coefficient, wps has a significant positive coefficient, and talk duration has a significant negative coefficient estimate.

Table 3. Basic Estimates of the Drivers of the Popularity of the Talk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables:	**LVPM**	**LVPM**	**LVPM**	**LVPM**	**LVPM**	**LVIEW**	**LVIEW**	**LVIEW**	**LVIEW**	**LVIEW**
Courageous	0.362**	0.454***	0.499***	0.0759	0.172	0.210**	0.206**	0.157	0.0952	0.180*
	(0.178)	(0.174)	(0.178)	(0.0908)	(0.105)	(0.0944)	(0.0956)	(0.0988)	(0.0945)	(0.107)
Fascinating	-0.166	-0.0791	-0.0153	0.239***	0.216**	0.310***	0.306***	0.239***	0.262***	0.228**
	(0.138)	(0.133)	(0.146)	(0.0811)	(0.0903)	(0.0850)	(0.0840)	(0.0886)	(0.0824)	(0.0931)
Funny	-0.115	-0.0856	0.00937	0.377***	0.385***	0.468***	0.467***	0.394***	0.394***	0.395***
	(0.164)	(0.154)	(0.164)	(0.0870)	(0.0971)	(0.0968)	(0.0964)	(0.0976)	(0.0889)	(0.101)
Informative	0.152	0.207	0.252*	-0.0396	-0.0155	0.0910	0.0883	0.0234	-0.0227	-0.0109
	(0.124)	(0.127)	(0.133)	(0.0780)	(0.0917)	(0.0805)	(0.0801)	(0.0825)	(0.0788)	(0.0926)
Ingenious	-0.686***	-0.709***	-0.648***	-0.133	-0.0781	-0.115	-0.114	-0.187	-0.113	-0.0699
	(0.176)	(0.173)	(0.185)	(0.116)	(0.140)	(0.124)	(0.124)	(0.129)	(0.122)	(0.145)
Inspiring	0.196*	0.273**	0.326**	0.114*	0.123	0.228***	0.224***	0.167**	0.128*	0.127
	(0.116)	(0.116)	(0.127)	(0.0658)	(0.0760)	(0.0772)	(0.0769)	(0.0790)	(0.0702)	(0.0802)
Persuasive	-0.849***	-0.649***	-0.591***	-0.168	-0.0490	-0.101	-0.111	-0.164	-0.142	-0.0440
	(0.197)	(0.193)	(0.201)	(0.104)	(0.119)	(0.120)	(0.120)	(0.124)	(0.107)	(0.123)
Time		-0.046***	-0.046***	0.008**	0.012***		0.002	0.001	0.009**	0.012***
		(0.0109)	(0.0108)	(0.0039)	(0.0045)		(0.0046)	(0.0045)	(0.0038)	(0.0044)
Wpm			-0.001	0.002***	0.001*			0.001**	0.002***	0.001*
			(0.00091)	(0.00056)	(0.00064)			(0.00059)	(0.00055)	(0.00063)
Wps			0.016*	-0.009***	-0.009**	-0.003	-0.008**	-0.009**	0.0164*	-0.009***
			(0.0084)	(0.0033)	(0.0036)	(0.0035)	(0.0033)	(0.004)	(0.0084)	(0.0033)
Year FE	no	no	no	yes	yes	no	no	no	yes	yes
Month FE	no	no	no	yes	yes	no	no	no	yes	yes
Location FE	no	no	no	no	yes	no	no	no	no	yes
Constant	10.20***	10.77***	10.61***	8.878***	9.429***	13.86***	13.84***	13.74***	14.02***	14.68***
	(0.171)	(0.226)	(0.307)	(0.224)	(0.406)	(0.0808)	(0.103)	(0.153)	(0.217)	(0.410)
Observations	2,130	2,130	2,130	2,130	2,130	2,130	2,130	2,130	2,130	2,130
R-squared	0.049	0.071	0.074	0.697	0.777	0.035	0.035	0.038	0.162	0.381

Note: The base category is "Beautiful". Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. FE stands for fixed effects.

From (3-1) to (3-3) there is no control for year, month, and location of talks. When only the year and month are controlled for, as in (3-4), the overall picture changes considerably. First, among the viewer ratings, "fascinating" (+) and "funny" (+) are added to significant regressors and "inspiring" maintains its positive coefficient where the other ratings remain insignificant. The effect of talk duration on LVPM turns to positive. More importantly, both wpm (+) and wps (-) become highly significant. Controlling further for talk locations, in (3-5), picture of (3-4) has almost been preserved except for this rating category "inspiring" losing its statistically significant coefficient estimate.

A similar sequence of equations has then been estimated by setting our dependent variable as LVIEW. Owing to a reasonable degree of similarities between this new sequence (3-6 to 3-10) and the previous one (3-1 to 3-5), here we suffice with a brief discussion of (3-10) where "courageous" (+), "fascinating" (+) and "funny" (+) have significant effects on LVPM, talk duration has a significant positive coefficient, and both wpm (+) and wps (-) are significant. In addition, across the models (3-8) to (3-10), the signs of time, wpm, and wps are all preserved. Still, one needs to include more direct indicators of rhetorical features and meaning before reaching better conclusions, as in the next subsection.

3.5. Incorporating the Rhetorical Features and Psychological Processes

In this section, we depart from our specifications (3-5) and (3-10) as they are significantly inclusive of all our basic set of variables and further incorporate the measures of rhetoric and meaning. In that, we consider three separate extensions of our basic specifications. First, we employ two groups of indicators generated by LIWC by processing the talk transcripts. The first is a bundle of summary variables, and the second comprises psychological process aggregates. This exercise is our central attempt to include rhetorical features and meaning. Second, we reconsider the designated focal words of Özmen and Yücel (2019). These exercises are separately elaborated below in that order.

3.5.1. LIWC-generated measures as explanatory variables

The analysis of the LIWC-generated measures stems from (3-5) and (3-10), reused as the first and fifth specifications in Table 4 for convenience. In (4-2), LIWC summary variables replace wpm and wps. Among them, analytic (-), tone (+) and authentic (+) turn out to be statistically significant, whereas clout (-) is not significant on popularity. Continuing further, in (4-3), LIWC's psychological process aggregates replace the summary variables. affect (+), social (+), cogproc (+), percept (+), bio (+) and drives (-) are statistically significant on LVPM, whereas relativ and informal seem not to impact LVPM. In (4-4), all indicators are pooled in a single regression. Across (4-1) through (4-4), talk duration (time) has a significantly positive coefficient estimate. The overall pattern in (4-1) to (4-4) is also revealed for (4-5, same as 3-10) to (4-8), where we take LVIEW as the dependent variable.

Table 4. Estimates when Rhetorical Features and Meaning Incorporated

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables:	**LVPM**	**LVPM**	**LVPM**	**LVPM**	**LVIEW**	**LVIEW**	**LVIEW**	**LVIEW**
Courageous	0.172	0.279***	0.245**	0.259**	0.180*	0.286***	0.257**	0.269**
	(0.105)	(0.102)	(0.106)	(0.114)	(0.107)	(0.102)	(0.107)	(0.114)
Fascinating	0.216**	0.266***	0.352***	0.331***	0.228**	0.279***	0.366***	0.344***
	(0.090)	(0.081)	(0.095)	(0.100)	(0.093)	(0.084)	(0.099)	(0.103)
Funny	0.385***	0.365***	0.334***	0.274**	0.395***	0.376***	0.345***	0.285**
	(0.097)	(0.098)	(0.106)	(0.108)	(0.101)	(0.102)	(0.110)	(0.112)
Informative	-0.016	0.055	0.111	0.111	-0.011	0.061	0.119	0.117
	(0.092)	(0.086)	(0.096)	(0.103)	(0.093)	(0.086)	(0.098)	(0.104)
Ingenious	-0.078	-0.030	0.113	0.101	-0.070	-0.020	0.127	0.114
	(0.140)	(0.131)	(0.152)	(0.155)	(0.145)	(0.137)	(0.158)	(0.161)
Inspiring	0.123	0.159**	0.207**	0.180**	0.127	0.165**	0.213**	0.186**
	(0.076)	(0.075)	(0.081)	(0.086)	(0.080)	(0.079)	(0.086)	(0.090)
Persuasive	-0.049	0.021	0.056	0.064	-0.044	0.026	0.063	0.069
	(0.119)	(0.122)	(0.126)	(0.131)	(0.123)	(0.125)	(0.130)	(0.135)

Table 4. Estimates when Rhetorical Features and Meaning Incorporated (Continue)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables:	**LVPM**	**LVPM**	**LVPM**	**LVPM**	**LVIEW**	**LVIEW**	**LVIEW**	**LVIEW**
time	0.012***	0.012***	0.011***	0.013***	0.012***	0.012***	0.011***	0.013***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
wpm	0.001*			0.001*	0.001*			0.001
	(0.001)			(0.001)	(0.001)			(0.001)
wps	-0.009**			-0.003	-0.009**			-0.003
	(0.004)			(0.003)	(0.004)			(0.003)
analytic		-0.006***		-0.003**		-0.006***		-0.003*
		(0.001)		(0.002)		(0.001)		(0.001)
clout		-0.002		-0.001		-0.002		-0.001
		(0.002)		(0.003)		(0.002)		(0.003)
authentic		0.001		0.005**		0.001		0.005**
		(0.001)		(0.003)		(0.001)		(0.003)
tone		0.003***		0.003***		0.003***		0.002***
		(0.001)		(0.001)		(0.001)		(0.001)
affect			0.077***	0.074***			0.076***	0.073***
			(0.016)	(0.016)			(0.016)	(0.016)
social			0.017**	0.020**			0.018**	0.021***
			(0.007)	(0.008)			(0.007)	(0.008)
cogproc			0.037***	0.011			0.039***	0.014
			(0.007)	(0.009)			(0.007)	(0.009)
perception			0.041***	0.035***			0.042***	0.037***
			(0.010)	(0.010)			(0.009)	(0.010)
Bio			0.036***	0.038***			0.036***	0.038***
			(0.013)	(0.013)			(0.013)	(0.013)
drives			-0.035***	-0.031***			-0.036***	-0.032***
			(0.012)	(0.012)			(0.011)	(0.011)
relativity			0.012	-0.014			0.011	-0.015
			(0.007)	(0.015)			(0.007)	(0.015)
informal			0.015	-0.034			0.012	-0.034
			(0.052)	(0.054)			(0.051)	(0.052)
Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Month FE	yes	yes	yes	yes	yes	yes	yes	yes
Location FE	yes	yes	yes	yes	yes	yes	yes	yes
Constant	9.429***	9.621***	8.417***	8.906***	14.678***	14.886***	13.652***	14.107***
	(0.406)	(0.430)	(0.388)	(0.495)	(0.410)	(0.441)	(0.402)	(0.510)
Observations	2,130	2,130	2,130	2,130	2,130	2,130	2,130	2,130
R-squared	0.777	0.782	0.789	0.792	0.381	0.396	0.417	0.426

Note: The base category for ratings is "Beautiful". Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. FE stands for fixed effects.

Although the findings for psychological process aggregates are enlightening, coefficient estimates for the aggregates may be biased. Taking affective processes as an aggregate, for instance, one considers positive emotion and negative emotion as a bundle, which is not much acceptable once we realize that negative emotion is further decomposable into anxiety, anger, and sadness. Similarly, the aggregate for drives comprises affiliation, achievement, power, reward, and risk; when relativity is taken, it encompasses all motion, space, and time. So, we may be bundling potentially opposite (as in the case of positive and negative emotions) or disjoint things (as in the case of motion, space, and time) together. As a resolution, we reconsider LIWC's psychological processes with particular attention to their sub-classes (or sub-dimensions) (Table 5).

Table 5. Estimates when Subclasses of Psychological Processes Incorporated

	LVPM	LVIEW		LVPM	LVIEW
Affective Processes (1)			*Core Drives and Needs*		
posemo	0.101***	0.101***	**affiliation**	-0.0239**	-0.0220**
negemo	0.0303	0.0307	**achieve**	-0.0144	-0.0164
Affective Processes (2)			**power**	-0.0542***	-0.0591***
posemo	0.100***	0.100***	**reward**	0.150***	0.149***
negemo			**risk**	0.0528*	0.0535*
anx	0.194***	0.183***	*Time Orientations*		
anger	-0.0877	-0.0915*	**focuspast**	-0.00145	-0.000950
sad	0.0505	0.0657	**focuspresent**	0.0121	0.0120
Social Processes			**focusfuture**	0.0540*	0.0462
family	0.0135	0.0236	*Relativity*		
friend	0.0260	0.0380	**motion**	0.00362	-0.00360
female	0.0148	0.00651	**space**	-0.0376***	-0.0393***
male	0.0130	0.0139	**timeliwc**	-0.00340	-0.00274
Cognitive Processes			*Personal Concerns*		
insight	0.0920***	0.0935***	**work**	-0.0119	-0.0111
cause	-0.0567**	-0.0595**	**leisure**	0.0149	0.0155
discrep	0.0146	0.0235	**home**	0.00442	0.00221
tentat	0.0631**	0.0615**	**money**	-0.0208	-0.0260
certain	-0.0118	-0.00447	**relig**	0.0127	0.0111
differ	0.00734	0.0110	*Informal Language*		
Perceptual Processes			**swear**	0.143	0.166
see	-0.00333	-0.00190	**netspeak**	-0.0468	-0.0692
hear	0.0574***	0.0590***	**assent**	0.253*	0.258*
feel	0.146**	0.141**	**nonflu**	-0.0441	-0.0385
Biological Processes			**filler**	-0.135	-0.119
body	0.0926***	0.0934***			
health	-0.0266	-0.0267			
sexual	0.143*	0.136*			
ingest	0.0538*	0.0541*			

Note: Robust standard errors are omitted to save space. *** p<0.01, ** p<0.05, * p<0.1. Each section of the table presents the results of a separate regression where the subclasses of a psychological process are considered. Each model also includes all the variables used in (3-5) and (3-10) as well, the output of which is omitted to save space.

Our findings from this exercise are as follows:

- Affective processes: *Positive emotions* significant positive, *negative emotions* positive but insignificant. When negative emotions are further decomposed into *anxiety*, *anger* and *sadness*, *positive emotions* maintain their significant positive coefficient, among the *negative emotions*, *anxiety* has a significant positive coefficient, and *anger* has a negative coefficient, which is significant only in $LVIEW$ equation. *Sadness* does not display any significance.
- Cognitive processes: *Insight* (+), *cause* (-) and *tentat* (+) have significant coefficient estimates, whereas *discrep*, *certain* and *differ* do not display any significance.
- Perceptual processes: *See* (-) is not significant; *hear* and *feel* both have positive impacts.
- Biological processes: *body*, *sexual* and *ingest* have significantly positive coefficient estimates; interestingly, *health* has a negative coefficient despite being insignificant.
- Drives: *Affiliation* and *power* have negative effects, both significant; *reward* and *risk* have significantly positive effects. *On the other hand, Achieve has an insignificant negative effec*t.
- Time orientations: Only *"focusfuture"* (+) has a significant effect and only for $LVPM$, where *"focuspast"* (-) and *"focuspresent"* (+) are not significant at all.

- Relativity: *Space* has a significant negative coefficient, whereas *motion* (mixed signs) and *time* (-) are both insignificant.
- Informal language: While *assent* has a significant positive coefficient, *swear* (+), *netspeak* (-), *nonflu* (-) and *filler* (-) do not display any significance.
- Social processes: No significance at all, with the coefficient signs of *family* (+), *friend* (+), *female* (+), *male* (+).
- Personal concerns: No significance at all, with the coefficient signs of *work* (-), *leisure* (+), *home* (+), *money* (-), *relig* (+).

Focusing on $LVPM$, we summarize the coefficient sign and significance information about psychological processes (aggregate (Table 4), disaggregate (Table 5)) in Table 6.

Table 6. Psychological Processes: Aggregate vs Disaggregate Effects on LVPM

Process	Impact	Process	Impact
Affective Processes (1): ++		***Drives: --***	
posemo	++	affiliation	--
negemo	+	achieve	-
Affective Processes (2): ++		power	--
posemo	++	reward	++
negemo		risk	++
anx	++	***Time Orientations: NA***	
anger	-	focuspast	-
sad	+	focuspresent	+
Social Processes: ++		focusfuture	++
family	+	***Relativity: +***	
friend	+	motion	+
female	+	space	--
male	+	time	-
Cognitive Processes: ++		***Personal Concerns: NA***	
insight	++	work	-
cause	--	leisure	+
discrep	+	home	+
tentat	++	money	-
certain	-	relig	+
differ	+	***Informal Language: +***	
Perceptual Processes: ++		swear	+
see	-	netspeak	-
hear	++	assent	++
feel	++	nonflu	-
Biological Processes: ++		filler	-
body	++		
health	-		
sexual	+		
ingest	+		

Note: The table summarizes the estimated effects presented in Table 5. The denoted signs refer to: ++: Positive significant, +: Positive insignificant, --: Negative significant, -: Negative insignificant coefficients.

3.5.2. Focal Words

In our second exercise, we analyze the possible impacts of a selection of the attention-driving words (focal words) used by Özmen and Yücel (2019). In Table 7, (7-1) through (7-3), we estimate models that include the viewer-assigned ratings, talk duration, and the focal words of Özmen and Yücel (2019), all controlled for year, month, and location of talks. We examine the inclusion of the focal words in the talk title (7-1, 7-4), in the talk transcript (7-2, 7-5), and in the talk title and transcript combined (7-3, 7-6).

Table 7. Estimates when the Focal Words of Özmen and Yücel (2019) are used

	(1)	(2)	(3)	(4)	(5)	(6)
	LVPM	LVPM	LVPM	LVIEW	LVIEW	LVIEW
Variables:	Title	Transcript	Title & Transcript	Title	Transcript	Title & Transcript
Courageous	0.235**	0.223**	0.221**	0.245**	0.235**	0.233**
	(0.0986)	(0.102)	(0.101)	(0.100)	(0.103)	(0.103)
Fascinating	0.264***	0.268***	0.265***	0.280***	0.283***	0.281***
	(0.0836)	(0.0834)	(0.0832)	(0.0875)	(0.0858)	(0.0854)
Funny	0.461***	0.380***	0.377***	0.474***	0.393***	0.390***
	(0.0977)	(0.102)	(0.103)	(0.101)	(0.103)	(0.104)
Informative	0.0497	0.0800	0.0786	0.0578	0.0908	0.0895
	(0.0853)	(0.0813)	(0.0816)	(0.0875)	(0.0823)	(0.0824)
Ingenious	-0.0121	0.0215	0.0194	-0.00128	0.0348	0.0328
	(0.133)	(0.130)	(0.130)	(0.140)	(0.135)	(0.136)
Inspiring	0.191***	0.209***	0.208***	0.198***	0.217***	0.216***
	(0.0699)	(0.0705)	(0.0704)	(0.0754)	(0.0756)	(0.0753)
Persuasive	0.0326	0.0448	0.0441	0.0435	0.0528	0.0523
	(0.117)	(0.110)	(0.110)	(0.121)	(0.114)	(0.114)
time	0.0131***	0.0121**	0.0122**	0.0132***	0.0126**	0.0127**
	(0.00475)	(0.00516)	(0.00518)	(0.00471)	(0.00520)	(0.00523)
“globe”	-0.182**	-0.134***	-0.135***	-0.208**	-0.135***	-0.136***
	(0.0795)	(0.0310)	(0.0309)	(0.0822)	(0.0305)	(0.0305)
“brain”	0.0708	0.138***	0.138***	0.0751	0.142***	0.142***
	(0.0757)	(0.0409)	(0.0408)	(0.0739)	(0.0412)	(0.0411)
"future”	-0.0436	0.00757	0.00477	-0.0442	-0.00224	-0.00453
	(0.0695)	(0.0338)	(0.0335)	(0.0711)	(0.0337)	(0.0335)
“child”	0.0112	-0.00500	-0.00500	0.0348	-0.00497	-0.00497
	(0.169)	(0.0336)	(0.0335)	(0.163)	(0.0344)	(0.0344)
“technology”	0.0272	-0.115***	-0.115***	0.0640	-0.112***	-0.112***
	(0.345)	(0.0367)	(0.0365)	(0.342)	(0.0366)	(0.0364)
“hope”	-0.163**	-0.0355	-0.0356	-0.0472	-0.0337	-0.0336
	(0.0645)	(0.0278)	(0.0280)	(0.0636)	(0.0279)	(0.0280)
“change”	-0.127*	-0.0154	-0.0173	-0.154**	-0.0189	-0.0209
	(0.0694)	(0.0328)	(0.0324)	(0.0626)	(0.0334)	(0.0332)
“magic”	0.508**	0.120**	0.124**	0.494**	0.108*	0.112**
	(0.219)	(0.0564)	(0.0568)	(0.226)	(0.0557)	(0.0562)
“myth”	0.388***	0.0862	0.0859	0.327**	0.102	0.102
	(0.119)	(0.0727)	(0.0719)	(0.159)	(0.0693)	(0.0687)
“math”	0.00717	0.0971*	0.0977*	-0.0452	0.101*	0.101*
	(0.106)	(0.0571)	(0.0570)	(0.128)	(0.0562)	(0.0560)
“science”	-0.139	0.0154	0.0112	-0.184*	0.00581	0.00197
	(0.112)	(0.0451)	(0.0453)	(0.0960)	(0.0451)	(0.0453)
“bad”	0.410	0.0576*	0.0577*	0.431	0.0586*	0.0588*
	(0.268)	(0.0309)	(0.0308)	(0.263)	(0.0310)	(0.0308)
“sex”	0.200*	0.0414	0.0412	0.208*	0.0356	0.0355
	(0.116)	(0.0428)	(0.0429)	(0.113)	(0.0417)	(0.0418)
Year FE	yes	yes	yes	yes	yes	yes
Month FE	yes	yes	yes	yes	yes	yes
Location FE	yes	yes	yes	yes	yes	yes
Constant	9.437***	9.481***	9.475***	14.69***	14.72***	14.72***
	(0.365)	(0.402)	(0.400)	(0.372)	(0.411)	(0.409)
Observations	2,130	2,130	2,130	2,130	2,130	2,130
R-squared	0.779	0.786	0.786	0.387	0.406	0.407

Notes: Base category is “Beautiful”. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Across these models, “courageous”, “fascinating”, “funny” and “inspiring” have robustly positive and significant coefficient estimates and talk duration (time) has a significant positive impact on LVPM. Focusing on the significant focal words only, we observe that:

- “Globe” is negative and significant in all.
- “Hope” is significantly negative in the title and loses its significance but not the sign when the transcript is considered.
- “Change” always has a negative coefficient but is significant only in the title.
- “Brain” is positive but significant only in the transcript and title & transcript cases.
- “Magic” has a significant positive coefficient in all three cases.
- “Myth” always has a positive coefficient but is significant only in the title.
- “Math” has a positive coefficient in all three cases but is significant when transcript is considered.
- “Bad” has a positive coefficient in all three cases but is significant when transcript is considered.
- “Sex” always has a positive coefficient but is significant only in the title.
- “Technology” is insignificant and positive in the title, yet significantly negative in transcript and title & transcript.
- “Future”, “child” and “science” do not display any statistical significance.

The pattern in (6-1) to (6-3) is revealed also for (6-4) to (6-6), where we take $LVIEW$ as the dependent variable.

4. Discussion

Our results, as provided in the previous section, suggest interesting insights into the features of the talks that feed the popularity. First, we revisit our findings regarding the talks' pace, complexity, and duration. Here, wpm and wps being indicative of the pace and complexity of talks, respectively, positive effect of wpm and negative effect of wps provide us with good insights. The popularity of talks seems to have increased at higher talk paces, i.e., people are more responsive to tempo and excitement. However, the average number of words per sentence turns out to be a deterring factor in some specifications. Historically, the audience of TED Talks has a preference to praise complex talks less.

Another important finding of the basic estimates crystalizes around the variable time, where its sign reverses upon introducing a basic set of year, month and location controls, and including wpm and wps. So, lengthy talks not necessarily damage popularity. On the contrary, once we install the necessary controls, talk length may be associated with higher popularity. Parallel to our findings, Kravvaris and Kermanidis (2014), who analyze educational videos on YouTube, find that longer duration and faster pace are among the characteristics of more popular short educational videos.

We must note that time, wpm and wps are merely mechanical and may only be far indicators of the rhetorical or meaning-related dimensions. In that, one may suppose a higher wpm value indicates higher excitement on the speaker's side, a lower wps value indicates higher tractability or longer time means potential boredom on the audience's side.

Analyzing the link between the talk ratings and popularity also reveals interesting points. In general, talks rated courageous, fascinating, funny, and inspiring are on average more popular than talks rated as beautiful, the baseline category. We may categorize these ratings according to the source of popularity to some extent. One might argue that talks rated courageous and inspiring generally refer to talks that establish emotional ties between the speaker and the audience, which could stem from the topic's sensitivity or the speaker's expression. Next, talks rated as fascinating may be linked to the quality of the scientific content of the talk. Meanwhile, talks rated funny reflects the entertainment value of the talks. Thus, we may argue that talks with interesting scientific content, entertaining, and arousing emotions are generally more popular than talks rated as beautiful.

Next, we discuss the main summary variables of the LIWC. The first one is analytical thinking (Analytic), which measures the extent to which the speaker uses words that build logical and hierarchical thought patterns. Our results show that less analytical talks are more popular, suggesting that talks with more narrative content enhanced with personal experiences attract more attention. Second, the Tone variable captures the emotional tone of the talk, which is a summary variable capturing both positive and negative emotions. The higher the value, the higher the positive tone of the talk. By construction, values above 50 indicate a more positive tone. As seen in Table 2, the mean value of the tone of the TED Talks in our sample is 52.75, revealing that, on average, the talks have a slightly positive tone. Our empirical results suggest that the higher the positive emotional tone, the more popular the talk is. Third, more authentic talks are also more popular. The other summary indicator, Clout (leadership, self-confidence, and social status expressed in talks) does not significantly impact the popularity of the talks. In a nutshell, we may argue that analytical and tone are more related to the context of the talk while clout is more linked with the personality or the speaker's style. In a recent study, Gheorghiu et al. (2020) analyze whether first impressions or evaluations generated in short time intervals affect the perceived scientific quality and the entertainment value of the talk. They find no evidence for the impact of such impressions, and they further reveal that the scientific quality and entertainment value evaluations of the TED Talks are independent of the gender, attractiveness, ethnic background, or age of the speaker.

As to the psychological processes captured by LIWC, we point to a potential divergence between the aggregated and disaggregated indicators. In that, we provide a more comprehensive treatment of LIWC than MacKrill et al. (2021). Based on the "conformity between aggregate and disaggregate estimates", Table 7 suggests that positive emotions and anxiety mainly drive the effect of affective processes on the popularity of talks; the effect of cognitive processes is driven by insight and tentativeness; the effect of perceptual processes is driven by hearing and feeling; the effect of biological processes is driven by body-related matters and the effect of drives is determined by affiliation and power. So, in a vast space of attributes related to psychological processes, a small portion of all seems to have attracted the audience to talks. This may be a meaningful input in understanding and interfering with individuals' attention filters. Some observed differences shed light on this. For instance, among negative affective processes, anxiety increases popularity, while anger and sadness do not affect popularity. Among cognitive processes, providing insights increases popularity but explaining the causes of issues reduces the popularity of the talks. Contents focusing on reward and risk are positively related to popularity, while contents related to affiliation and power are negatively associated with popularity. Although the aggregate time orientation is not associated with popularity, focusing on the future significantly increases the view counts of the talks, pointing to another critical aspect of user preferences.

Several previous studies try to predict the ratings or emotional tone of the TED Talks. Bertero and Fung (2017) test different algorithms for speech emotion detection on a sample of TED Talks and find that algorithms are better at detecting "angry" and "sad" than "happy" as emotions of the speech. Tanveer et al. (2018) investigate the relation between narrative trajectories and the ratings of the talks by decomposing the affective components of the talk at various intervals. They show that talks with different ratings have different trajectories to some extent and suggest that varying the emotions during the talk, building a great ending, and initiating a snowball effect are key elements for a successful talk. In a similar study, Cullen and Harte (2017) investigate pieces of the talk to predict inspiring, funny and persuasive user ratings. They find that longer slices and slices towards the end of the talk have greater power to predict the rating of the talk. Unlike these studies, we directly focus on popularity taking a wider range of context and content wise indicators in the current study.

As a closing remark in our discussion, we must admit that this study assumes that the people (online audience) have found their way to the TED Talks purposefully and in the absence of any assistance. Nevertheless, there are recommender systems and several lines of assistance to direct people to TED Talks material. Indeed, the very "watch next" suggestions of the TED portal are of that kind. So, an implicit technical assumption of this study is that the recommender systems follow the behavioral patterns revealed in the study rather than altering them. In that, omitting recommender systems simply arises from its limited contribution to the current research focus, and TED's own recommender routines lie in our future research agenda.

5. Conclusion

This paper aims to study how people utilize (search for, choose, process, and evaluate) information provided on online domains, emphasizing the balance between context identifiers and the actual content of information and the psychological processes. To this end, the study assesses the popularity of online provided materials -TED Talks- in relation to the length of information, user ratings, and several content-related features, including psychological processes.

Our findings offer implications for the economics of online information provision specifically for online content suuplier seeking to maximize their outreach. Our analysis reveals that not only the content but also the contextual features of the information can significantly increase the view counts. Therefore, for instance, the title of the information should be appealing to the audience, and the length and pace of the information should be carefully curated. On the context side, focusing on the future, stimulating positive emotions, providing insights, emphasizing risks and rewards, and being less analytical are deemed to increase the popularity of the information/content.

Given that increased popularity is directly linked with extensive outreach on the Internet, the information providers can extend their success by thoroughly considering the balance of content and context, as well as the associated psychological processes.

Declarations and Disclosures

Ethical Responsibilities of Authors: The authors of this article confirm that their work complies with the principles of research and publication ethics.

Conflicts of Interest: No potential conflict of interest was reported by the authors.

Funding: The authors received no financial support for the preparation and/or publication of this article.

Author Contributions: The authors confirm contribution to the article as follows: Conceptualization and design, M. U. Özmen and M. E. Yücel; data collection, M. U. Özmen and M. E. Yücel; analysis of data and interpretation of results, M. U. Özmen and M. E. Yücel; writing the first draft of the manuscript, M. U. Özmen and M. E. Yücel; review and editing, M. U. Özmen and M. E. Yücel. The manuscript/article was read and approved by all the authors, and all authors accepted responsibility for their article.

Plagiarism Checking: This article was screened for potential plagiarism using a plagiarism screening program.

References

Almaged, S. (2021). Disseminating knowledge: A discourse analysis of terrorism in TED Talks. *Heliyon*, *7*(2), e06312. https://doi.org/10.1016/j.heliyon.2021.e06312

Azzopardi, L. (2021). Cognitive biases in search: A review and reflection of cognitive biases in information retrieval. Presented in Session 1: Cognitive Aspects of CHIIR '21, Canberra, Australia, March 14-19, 2021.

Bertero, D., & Fung, P. (2017). A first look into a convolutional neural network for speech emotion detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 5115-5119.

Broadbent, D. (1958). *Perception and communication*. London: Pergamon Press.

Budzinski, O., & Gaenssle, S. (2018). The economics of social media (super-) stars: An empirical investigation of stardom and success on YouTube. *Journal of Media Economics, 31*(3-4), 75-95.

Cettolo, M., Girardi, C., & Federico, M. (2012). WIT3: Web inventory of transcribed and translated talks. *Proceedings of the 16th EAMT Conference, 28-30 May 2012, Trento, Italy.*

Chen, L., & Lee, C. M. (2017). Predicting audience's laughter using convolutional neural network. arXiv preprint arXiv:1702.02584.

Cullen, A., & Harte, N. (2017). Thin slicing to predict viewer impressions of TED Talks. In *The 14th International Conference on Auditory-Visual Speech Processing*, 58-63.

Deutsch, J., & Deutsch, D. (1963). Attention: some theoretical considerations. *Psychological Review, 70*(1), 80-90.

Di Carlo, G. S. (2014). The role of proximity in online popularizations: The case of TED Talks. *Discourse Studies, 16*(5), 591-606.

Fischer, O., Jeitziner, L. T., & Wul, D. U. (2021). Affect in science communication: A data-driven analysis of TED Talks, PsyArXiv. https://doi.org/10.31234/osf.io/28yc5

Fiksdal, A. S., Kumbamu, A., Jadhav, A. S., Cocos, C., Nelsen, L. A., Pathak, J., & McCormick, J. B. (2014). Evaluating the process of online health information searching: A qualitative approach to exploring consumer perspectives. *Journal of Medical Internet Research, 16*(10), e224.

Gao, L., & Bai, X. (2014). Online consumer behaviour and its relationship to website atmospheric induced flow: Insights into online travel agencies in China. *Journal of Retailing and Consumer Services, 21*(4), 653-665.

Gheorghiu, A. I., Callan, M. J., & Skylark, W. J. (2020). A thin slice of science communication: Are people's evaluations of TED Talks predicted by superficial impressions of the speakers? *Social Psychological and Personality Science, 11*(1), 117-125.

Gordon, I. D., Chaves, D., Dearborn, D., Hendrikx, S., Hutchinson, R., Popovich, C., & White, M. (2022). Information seeking behaviors, attitudes, and choices of academic physicists. *Science & Technology Libraries*, *41*(3), 288-318.

Hammond, C. (2012). *Time warped: Unlocking the mysteries of time perception.* Edinburgh and London: Canongate.

Huberman, B. A. (2009). Social attention in the age of the web. In *Working Together or Apart: Promoting the Next Generation of Digital Scholarship* (pp. 62-69). Council on Library and Information Resources.

Hyland, K. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes, 9*, 116-127.

Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice-Hall, Englewood Cliffs.

Khan, G. F., & Vong, S. (2014). Virality over YouTube: An empirical analysis. *Internet Research, 24*(5), 629-647.

Kinnaird, K. M., & Laudun, J. (2019). TED talks as data. *Journal of Cultural Analytics,* 4(2), 1-21.

Kravvaris, D., & Kermanidis, K. L. (2014). Speakers' language characteristics analysis of online educational videos. In *IFIP International Conference on Artificial Intelligence Applications and Innovations,* 60-69. Springer, Berlin, Heidelberg.

Lanham, R. A. (2006). *The economics of attention: Style and substance in the age of information*. Chicago: University of Chicago Press.

Lee, L. H., Juan, Y. C., Tseng, W. L., Chen, H. H., & Tseng, Y. H. (2015). Mining browsing behaviors for objectionable content filtering. *Journal of the Association for Information Science and Technology, 66*(5), 930-942.

Liao, S. H., Widowati, R., & Hsieh, Y. C. (2021). Investigating online social media users' behaviors for social commerce recommendations. *Technology in Society, 66,* 101655.

Lim, J. S., Nicholson, J., Yang, S. U., & Kim, H. K. (2015). Online authenticity, popularity, and the "Real Me" in a microblogging environment. *Computers in Human Behavior, 52*, 132-143.

Liou, H. C., & Tseng, Y. P. (2022). A corpus-based study of phrasal verbs with key meanings in TED Talks. *English Teaching & Learning, 46*, 395-415.

Liu, Z., Xu, A., Zhang, M., Mahmud, J., & Sinha, V. (2017). Fostering user engagement: Rhetorical devices for applause generation learnt from TED Talks. In *Proceedings of the International AAAI Conference on Web and Social Media, 11*(1).

Lopes, J., Trancoso, I., & Abad, A. (2011). A nativeness classifier for TED Talks. *Acoustics, Speech and Signal Processing (ICASSP), 2011* IEEE International Conference on, 22-27 May 2011. https://doi.org/10.1109/ICASSP.2011.5947647

Maeno, S., & Maeshiro, T. (2018). Prediction of standing ovation of TED Technology Talks. In *International Conference on Human Interface and the Management of Information,* 677-684. Springer Cham.

MacKrill, K., Silvester, C., Pennebaker, J. W., & Petrie, K. J. (2021). What makes an idea worth spreading? Language markers of popularity in TED Talks by academics and other speakers. *Journal of the Association for Information Science and Technology, 72*(8), 1028-1038.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behaviour.* New York, NY: Holt, Rinehart & Winston.

Mou, J., & Shin, D. (2018). Effects of social popularity and time scarcity on online consumer behaviour regarding smart healthcare products: An eye-tracking approach. *Computers in Human Behavior, 78*, 74-89.

Norman, D. (1968). Toward a theory of memory and attention. *Psychological Review, 75*(6), 522-536.

Özmen, M. U. (2015). Online information retrieval behaviour and economics of attention. *Online Information Review*, *39*(6), 779-794.

Özmen, M. U. & Yücel, M. E. (2019). Handling of online information by users: Evidence from TED talks. *Behaviour & Information Technology*, *38*(12), 1309-1323.

Pappas, N., & Popescu-Belis, A. (2013). Sentiment analysis of used comments for one-class collaborative filtering over TED Talks. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 773-776. https://doi.org/10.1145/2484028.2484116

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC 2015*. Austin, TX: University of Texas at Austin. https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf (Access Date: 12 October, 2022).

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015. Austin*, TX: Pennebaker Conglomerates. https://www.researchgate.net/publication/337731895_Linguistic_Inquiry_and_Word_Count_LIWC2015 (Access Date: 12 October, 2022).

Pozdena, E. (2020). Cultural dimensions and characteristics reflected in Hungarian TED Talks. *The European Journal of Humour Research, 8*(2), 62-67.

Reisoğlu, İ., Çebi, A., & Bahçekapılı, T. (2022) Online information searching behaviours: Examining the impact of task complexity, information searching experience, and cognitive style. *Interactive Learning Environments, 30*(3), 417-434.

Romanelli, F., Cain, J., & McNamara, P. J. (2014). Should TED Talks be teaching us something. *American Journal of Pharmaceutical Education, 78*(6), 113.

Rousseau, A., Deleglise, P., & Esteve, Y. (2014). Enhancing the TED-LIUM corpus with selected data for language modeling and more TED Talks. *Language Resources and Evaluation Conference (LREC 2014)*, Iceland.

Rubenstein, L. D. (2012). Using TED Talks to inspire thoughtful practice. *The Teacher Educator, 47*(4), 261-267.

Schwemmer, C., & Jungkunz, S. (2019). Whose ideas are worth spreading? The representation of women and ethnic groups in TED Talks. *Political Research Exchange, 1*(1), 1-23.

Segev, E., & Ahituv, N. (2010). Popular searches in Google and Yahoo!: A "digital divide" in information uses? *The Information Society, 26*(1), 17-37.

Simola, J., Hyönä, J., & Kuisma, J. (2014). Perception of visual advertising in different media: From attention to distraction, persuasion, preference and memory. *Frontiers in Psychology, 5,* 1208.

Simon, H. A. (1971). Designing organizations for an information-rich world. In Greenberger, M. (Ed.), *Computers, communication, and the public interest.* Baltimore, MD: The Johns Hopkins Press.

Skarpa, P. E., & Garoufallou, E. (2021). Information seeking behavior and COVID-19 pandemic: A snapshot of young, middle aged and senior individuals in Greece. *International Journal of Medical Informatics, 150,* 104465.

Sugimoto, C. R., Thelwall, M., Larivie`re, V., Tsou, A., Mongeon, P., & Macaluso, B. (2013). Scientists popularizing science: Characteristics and impact of TED talk presenters. *PLoS ONE, 8*(4), e62403.

Tanveer, M. I., Samrose, S., Baten, R. A., & Hoque, M. E. (2018). Awe the audience: How the narrative trajectories affect audience perception in public speaking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-12.

Tafesse, W. (2020). YouTube marketing: How marketers' video optimization practices influence video views. *Internet Research, 30*(6), 1689-1707.

Treisman, A. (1964). Selective attention in man. *British Medical Bulletin, 20*(1), 12-16.

Tsou, A., Thelwall, M., Mongeon, P., & Sugimoto, C. R. (2014). A community of curious souls: An analysis of commenting behavior on TED Talks videos. *PLoS ONE, 9*(4), e93609.

Tur, B., Harstad, J., & Antonakis, J. (2018). Effect of charisma in informal leadership settings: The cases of TED and Twitter. In *Academy of Management Proceedings, 2018*(1), 13242. Briarcliff Manor, NY: Academy of Management.

Wingrove, P. (2022). Academic lexical coverage in TED Talks and academic lectures. *English for Specific Purposes, 65,* 79-94.

Wook, T. S. M. T., & Salim, S. S. (2014). Visual interface for searching and browsing children's WebOPAC. *Universal Access in the Information Society, 13*(4), 367-385.

Appendix

Table A1. The Descriptive Statistics of all LIWC Dimensions

LIWC Dimension	Label	Mean	Stdev	LIWC Dimension	Label	Mean	Stdev
Word Count	*WC*	2084.40	772.44	**Perpetual Processes**	*percept*	2.88	2.15
Summary Variables				Seeing	*see*	1.42	1.92
Analytical Thinking	*Analytic*	63.20	14.89	Hearing	*hear*	0.91	1.19
Clout	*Clout*	80.40	9.80	Feeling	*feel*	0.38	0.34
Authentic	*Authentic*	28.77	13.23	**Biological Processes**	*bio*	1.74	1.38
Emotional Tone	*Tone*	52.75	20.15	Body	*body*	0.56	0.66
Language Metrics				Health/illness	*health*	0.74	0.85
Words per sentence	*WPS*	16.72	4.30	Sexuality	*sexual*	0.09	0.26
Words>6 letters	*Sixltr*	17.87	3.59	Ingesting	*ingest*	0.33	0.56
Dictionary words	*Dic*	82.59	4.57	**Core Drives and Needs**	*drives*	8.09	2.30
Function Words	*function*	51.39	5.75	Affiliation	*affiliation*	3.31	2.05
Total pronouns	*pronoun*	13.27	2.77	Achievement	*achieve*	1.46	0.71
Personal pronouns	*ppron*	6.51	1.87	Power	*power*	2.31	1.02
1st pers sing.	*i*	0.78	0.79	Reward focus	*reward*	1.16	0.51
1st pers plur.	*we*	2.16	1.23	Risk/prevention focus	*risk*	0.48	0.38
2nd person	*you*	1.71	1.04	**Time Orientation**			
3rd pers sing.	*shehe*	0.72	0.92	Past focus	*focuspast*	3.74	1.79
3rd pers plur.	*they*	1.13	0.70	Present focus	*focuspresent*	10.52	2.43
Impersonal pronouns	*ipron*	6.76	1.78	Future focus	*focusfuture*	1.03	0.49
Articles	*article*	7.33	1.72	**Relativity**	*relativ*	13.28	2.31
Prepositions	*prep*	12.87	2.11	Motion	*motion*	2.00	0.75
Auxiliary verbs	*auxverb*	8.16	1.64	Space	*space*	6.97	1.79
Common adverbs	*adverb*	5.67	1.42	Time	*time*	4.49	1.82
Conjunctions	*conj*	7.01	1.50	**Personal Concerns**			
Negations	*negate*	1.23	0.53	Work	*work*	2.52	1.54
Grammar Other				Leisure	*leisure*	0.88	1.26
Regular verbs	*verb*	15.26	2.83	Home	*home*	0.28	0.32
Adjectives	*adj*	4.28	1.71	Money	*money*	0.65	0.87
Comparatives	*compare*	2.26	0.74	Religion	*relig*	0.18	0.40
Interrogatives	*interrog*	1.89	0.62	Death	*death*	0.18	0.33
Numbers	*number*	3.92	3.09	**Informal Speech**	*informal*	0.42	0.40
Quantifiers	*quant*	2.31	0.71	Swear words	*swear*	0.03	0.07
Affect Words	*affect*	3.99	1.46	Netspeak	*netspeak*	0.06	0.17
Positive emotion	*posemo*	2.70	1.07	Assent	*assent*	0.14	0.21
Negative emotion	*negemo*	1.23	0.83	Nonfluencies	*nonfl*	0.18	0.17
Anxiety	*anx*	0.23	0.28	Fillers	*filler*	0.01	0.05
Anger	*anger*	0.31	0.40	**All Punctuation**	*allpunc*	19.20	5.20
Sadness	*sad*	0.23	0.23	Periods	*period*	5.66	1.38
Social Words	*social*	10.44	3.84	Commas	*comma*	6.61	1.56
Family	*family*	0.28	0.44	Colons	*colon*	0.30	0.47
Friends	*friend*	0.16	0.19	Semicolons	*semic*	0.08	0.12
Female referents	*female*	0.48	0.83	Question marks	*qmark*	0.49	0.38
Male referents	*male*	0.70	0.88	Exclamation marks	*exclam*	0.04	0.13
Cognitive Processes	*cogproc*	11.21	2.54	Dashes	*dash*	1.64	1.02
Insight	*insight*	2.42	0.91	Quotation marks	*quote*	0.77	0.82
Cause	*cause*	1.97	0.72	Apostrophes	*apostro*	2.58	1.10
Discrepancies	*discrep*	1.36	0.55	Parentheses (pairs)	*parenth*	0.94	3.53
Tentativeness	*tentat*	2.36	0.84	Other punctuation	*otherp*	0.08	0.38
Certainty	*certain*	1.35	0.49				
Differentiation	*differ*	2.95	0.87				

Notes: The number of observations is 2130 (6 to 20 minute-long talks are considered). More about LIWC: https://liwc.wpengine.com