

Handling of online information by users: evidence from TED talks

M. Utku Özmen & Eray Yucel

To cite this article: M. Utku Özmen & Eray Yucel (2019) Handling of online information by users: evidence from TED talks, Behaviour & Information Technology, 38:12, 1309-1323, DOI: [10.1080/0144929X.2019.1584244](https://doi.org/10.1080/0144929X.2019.1584244)

To link to this article: <https://doi.org/10.1080/0144929X.2019.1584244>



Published online: 27 Feb 2019.



Submit your article to this journal [↗](#)



Article views: 193



View related articles [↗](#)



View Crossmark data [↗](#)



Handling of online information by users: evidence from TED talks

M. Utku Özmen^a and Eray Yucel^b

^aResearch and Monetary Policy Department, Central Bank of the Republic of Turkey, Ankara, Turkey; ^bDepartment of Economics, Ihsan Dogramaci Bilkent University, Ankara, Turkey

ABSTRACT

This paper studies how people search for, choose, process and evaluate information provided online. In this context, the study analyses how the content and context of online information are related to the length of information and to user ratings. Employing naturalistic data that cover the titles, durations and viewer-assigned ratings/tags of more than two-thousand TED talks, the paper investigates whether (i) the talk duration is related to viewer-assigned ratings, (ii) there is a link between the talk duration and attention driving factors (title words), and (iii) the ex-ante wording of talks' titles and ex-post user-assigned ratings are connected. The findings show that talks with certain end-user ratings have significantly different length, most strikingly, talks first rated as persuasive are on average 35% longer than talks first rated as ingenious. Also the inclusion of certain words in the talk title significantly affects both the talk duration and end-user ratings. For instance, talks whose title include 'child' are on average 27% longer than other talks; or talks whose title include 'brain' are 57% more likely to be rated as fascinating than others. Overall, the paper reveals regularities regarding information processing attitudes, attention and subjective evaluations of online information users.

ARTICLE HISTORY

Received 14 June 2017
Accepted 10 February 2019

KEYWORDS

TED talks; attention; online viewing behaviour; information processing; Behavioural sciences

A wealth of information creates a poverty of attention.
(Herbert A. Simon)

1. Introduction

Increased availability of online information resources, information technology and the emergence of new online social domains have changed dramatically the way people communicate, search for information and disseminate ideas. Such a structural change in the accessibility and abundance of information is also challenging the classical methods through which people know things, re-introducing a dilemma between the content and the context of information. A concrete example may be followed from an interesting talk TEDxNewYork aired at mid-January of 2015. The presenter, Will Stephen, entitled his talk 'How to sound smart in your TEDs talk?', and the talk was classified as comedy by its publishers.¹ As the authors of this paper, neither of whom viewed the talk on-site, we enjoyed the talk with appreciation simply because it underlined an important problem via the vivid power of comedy. At the cost of giving a spoiler, Stephen (2015) based his satire on a hypothetical speaker who is well-marketing 'nothing' as 'something worthy to listen'. In that, something not smart, not

inspirational or even not remotely researched can be well-transmitted to others as something brilliant with some good sense of learning if the speaker maintains a specific mannerism during the talk. In Stephen's way of caricaturising, this might include some hand gestures, a theatrical use of accessories like glasses, posing a sharp question at the entrance, then telling an anecdote to break the audience's tension as well as to buy some time, then to dwell on the theme of the talk via a sequence of numerical figures, statistics, charts even irrelevant, attracting words accompanied by 'vaguely thought provoking stock images' up until whatever has been presented builds to a moment which is the climax of the talk, possibly coming right before the talk ends. When taken literally, this caricature would be too unfair of a criticism. However, not taking it seriously at all would be unfair to the speaker, Mr. Stephen, as well.

To us, the climax of the talk was when Stephen holds his glasses in the air and says that there were not even glasses fixed in the frames, highlighting the inherent duality between the content and context. Is the best gift under the Christmas tree in the shiniest box, or not? This is not a new question, indeed, and has occurred continually during the intellectual history of humans. 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. Classical mechanisms of research, writing, face-to-face sharing of views, critiques, debates and the like served well to locate that fine line practically until the 1990s. During these classical ages of information seeking, cost structure of producing and using knowledge favoured intellectual quality; needless to say, a relatively smaller portion of people in each society had access to domain of intellectual products then. By the 1990s, the world underwent a massive change due to expansion of Internet at dramatic rates. Generations of World Wide Web, global computer hardware stock evolved in a way to provide a full (and even widening) spectrum of 'information' services in the present. The quotation marks are intentional here, since it can no more be assured that people reach, acquire, digest and share knowledge. How human beings know things has transformed, a larger portion of societies has begun accessing and sharing information with the help of technology, where key parameters in the economics of attention changed drastically and irreversibly. Our times, then, may be defined as a new age of knowing things with a different mannerism and code of communication.

Formation of a full network of selecting, processing, re-processing and disseminating information can be attributed to almost zeroed marginal costs associated with these processes. Concurrently, increased media diversity and a more democratic view of media access mandated a re-invented and improved version of traditional library search schemes, resulting in a state of the art page ranking algorithms some decades after Google's first-page rank algorithm. So, among the many factors, availability of online content can be seen as something revolutionary to cut down individuals' everyday research budgets, in terms of time and physical as well as intellectual labour by bending the domain of perception and (selective) attention. As noted in Özmen (2015), since online information retrieval is free money-wise, amount of attention is the price paid for knowing things.

Following a similar path, we first analyse how the length of an information stream is related to some user-assigned rankings attached to it. In the current context, the information streams are talks aired via Internet and the user-assigned rankings are a list of adjectives that users assign to talks upon viewing them. Second, we expand the exercise so as to reveal possible linkages between the length of information streams, i.e. talks, and some specific attention driving words appearing in talks' titles. In this way, one *ex post* variable (user-assigned rankings) and one *ex ante* variable (wording of talks' titles) are connected to length of information streams (length of talks). This exercise gives us the

opportunity to study the information selecting and processing attitudes of individuals in the context of online viewing behaviour. Selection of information corresponds to choosing of talks among an available set, which may be induced by keywords, the title, ratings or the appearance of the speaker. Processing of information refers to how the information is internalised and what has been taken out from the information, which to some extent may be traced in the subjective evaluations and ratings of the viewers. Information choice and processing attitudes may indeed be different on online domains than in physical domains. For instance, in our case, the audience who attended the talks had a smaller set of information which was limited to the talk title and appearance of the speaker. On the other hand, online viewers of the talks access a larger set including additional information such as the comments and ratings of previous viewers or the viewing suggestions of the website itself, which may affect their viewing behaviour. Thus, answers to abovementioned questions are especially important for the online content providers in improving their collections or to expand their outreach.

Using the titles, lengths and viewer-assigned ratings of more than two-thousand TED talks, we try to reach some statistical conclusions on information processing attitudes and online viewing behaviour of individuals. The difference of TED's talk archive from online photo galleries, online grocery sites, etc. is that the viewers cannot have a good sense of the content without viewing talk videos fully. Furthermore, assuming that they completely watch every talk they clicked, viewers cannot advance among talks in less than 6 minutes or in more than (mostly) 18 minutes. So, the cost of consumption (watching a talk) is still non-negligible here; yet the cost of observing package information is virtually zero. Consequently, our prior takings were such that there will be talks with durations significantly less than 18 minutes, accounting for the impatience in contemporary societies. However, we did not expect any linkage between viewer-assigned ratings and length of talks. Moreover, with some anticipation that TED viewers would be more content – rather than title-oriented, we did not expect any strong connection between viewer-assigned ratings and titles of talks either. These anticipations locate our point of departure and define our research question in broad terms. In the end, we reveal statistically significant linkages between talk durations and viewer assigned ratings as well as talk durations and certain wording of talk titles. Relations between viewer-assigned ratings and wording of talk titles are also of an interesting nature.

The findings have certain implications for online content and technology providers in improving their

outreach. The amount (length) of online information consumed by people is affected by the attractiveness of the presentation and both content and context alter people's attention. Also, from the context side, certain attitudes of the information provider (i.e. presentation style, selection of topics) as well as textual information (i.e. title of information) have an impact on the subjective evaluations of the end-users of information, in terms of ratings awarded. Given that certain tags attached to a piece of information by users may further attract other users, online information providers may benefit from the findings of this study when designing the way they provide information in a more targeted manner.

The remainder of the study is organised as follows: in the next section, background of the current research is given. Section 3 presents our data and lays down the empirical strategy, the results of which are provided in Section 4. Section 5 further discusses the findings and concludes the paper.

2. Background

2.1. Cognitive psychological foundations of information processing

Cognitive psychology provides us with a number of explanations of attention, i.e. selective concentration. Facing a large number of stimuli, the individual filters the unwanted ones in a way to minimise her subsequent cognitive effort. Broadbent's (1958) bottleneck theory stating that excessive information that cannot be handled by one are simply ignored and Treisman's (1964) modification of Broadbent stating that at the early stage, available set of stimuli is processed in a parallel manner, while the selection is made 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 studies, pertinence of information induces filtering and selection to occur at a later stage; calling for an active processing strategy defined by person's goals. In an information-abundant environment as experienced today, selective attention refers to attending to information that maximise utility with respect to some goals. In that, Miller, Galanter, and Pribram's (1960) information processing theory defines the test-operate-test-exit as the basic unit of behaviour. Information is processed in a sequential manner 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), availability of too much information results in poverty of attention, suggesting a need for efficient allocation of attention. Kahneman

(1973) points at a possible upper limit for resources, including attention, that a person devotes 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 under a different light welcoming the role of information providers as advertisers. Advertising, then, is increasing the quality perceived by the users through initial stimuli as discussed by Huberman (2009) and Simola, Hyönä, and Kuisma (2014). The interested reader might also visit Kahneman and Tversky (1979) for the role of uncertainty in decision-making instances.

2.2. Earlier work considering the user attitudes toward online viewing behaviour

The literature is not dry as far as the web sites and online libraries of still images are concerned. How users' choice and browsing behaviour are determined by the characteristics of online information and how the online viewing behaviour is affected through online images and digital photos have all been studied well. In that, Zhang et al. (2014) find that content factors are more important than contextual factors, Fiksdal et al. (2014) report information saturation and fatigue as main reasons for stopping information retrieval and Wook and Salim (2014) identify specification requirements for visual aspects of information provision as the use of space, organisation of information, and function and use of colour. Moving further, predictive models of web browsing behaviour based on its past records are viable as suggested by Lee et al. (2015). Importance of the information provision style and consumers' perception is reported by Hsieh et al. (2015) and Gao and Bai (2014), confirm like other studies above the cognitive processing of online information by visitors. The online information viewing attitudes of users may also differ according to cultural background and the organisation of societies. As pointed out by Segev and Ahituv (2010), regarding textual context of search terms, while in some countries internet searches are more motivated by socio-political concerns, in others searches are more motivated by entertainment concerns. Özmen (2015) provides a good account of the studies of online collections of still images, underlining the attractiveness of photos, tags, dominance of visual content over the textual as well as the attention-augmenting role of photos.

2.3. Treatment of TED talks in the recent literature

The domain of TED talks has nicely provided a fruitful venue for many researchers. In that, Lopes, Trancoso, and Abad (2011) study the acoustic- and speech recognition aspects, Rousseau, Deleglise, and Esteve (2014) use TED data in language modelling and Cettolo, Girardi, and Federico (2012) build a Web inventory of transcribed and translated talks. In our case, we focus on attention economic dimensions which recently gained some importance. Among the recent studies, Tsou et al. (2014) examine the audience reactions to understand the impacts of presenter characteristics and platform on the reception of a video. By means of a content analysis of comments left on both the TED's website and the YouTube platform, they find that commenters were more likely to discuss the characteristics of a presenter on YouTube, whereas commenters tended to engage with the talk content on the TED website. Furthermore, people tended to be more emotional, positively or negatively, when the speaker was a woman. Pappas and Popescu-Belis (2013) also focus on user comments not accompanied by explicit rating labels and investigate their utility for a one-class collaborative filtering task such as bookmarking, where only the user actions are given as ground truth.

As a venue to disseminate scientific knowledge among others, TED talks are studied by Sugimoto et al. (2013) in an attempt to reveal its publicity-enhancing role within academia. They find that giving a TED presentation has no association with the number of citations subsequently received by an academic. TED as a populariser of research, might not promote the work of scientists within the academic community. Based on Sugimoto et al. (2013), it seems TED does not replace the still-conventional channels of academic communication. Despite, its educational roles or possible contribution to educators are not to be dismissed directly. As discussed by Romanelli, Cain, and McNamara (2014), TED stepped up as an institutional star of popular culture in 2006, once the curators began providing short, free, unrestricted and educational videos. Note that the brand name of TED dates back to 1984 and one may attribute its spread at an epidemic rate not only to ease in access but also to some 22 years of established reputation. In their assessment of the role of TED as a venue of teaching, Romanelli, Cain, and McNamara (2014) assert that TED talks seem to accomplish the goals of spreading ideas while sparking curiosity within the learner and they question whether the academia can benefit from a potential nexus of classical classroom and this new 'style' of communication. Rubenstein (2012) can also

be viewed with regard to TED talks' relevant content that informs teachers of best practices, current issues, and innovative future possibilities.

Downsides of TED with regard to its learning-related functions are also underlined by Romanelli, Cain, and McNamara (2014). 'Flattening or dumbing down ideas', 'being primarily designed to entertain' and 'generating a false sense of simplicity' can be counted in here. As a consequence, TED-like environments might suppress the learning efforts which might otherwise be more directed toward better academic conceptualisations.

Di Carlo (2014) approaches the TED talks via Hyland's concept of proximity along with its five elements, namely organisation, argument structure, credibility, stance and reader engagement. They say the talks emphasise proximity of commitment, by concentrating not on the speakers' identity and reputation, but rather on how they are personally involved in the topic of the speech; revealing TED's idea that science should be ideas to be discussed rather than information to be passively received.

All in all, given our research question and the earlier literature of the economics of attention as well as cognitive psychology, the Internet-based collection of TED talks seems to be a suitable empirical environment. Despite it lacks certain characteristics of a full-fledged experimental setup, the empirical strategy we elaborate in the next section allows us to extract quite a reliable body of information from raw data.

3. Data and empirical strategy

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, www.ted.com. The dataset covers all 2222 talks posted starting from June 2006 (earliest) up to first half of June 2016. The data are retrieved from the general page where all the talks are listed together in subsequent sub-pages, (i.e. <https://www.ted.com/talks?page=1>). Such a page includes a small box for each talk with a photo-link which directs the viewer to the video page of the talk. Under the picture, name of the presenter and title of the talk, date of posting are located along with ratings. Thus, we collect all the following information from these general pages: the name of the presenter, 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.² In that sense, the way that the data is organised on the website is an example of an attribute-based information

provision, where title, duration and ratings relate to certain attributes of the talk to be chosen to be viewed.

From a statistical standpoint, our data set is a naturalistic one as we use data from all users in a real setting. Moreover, as long as the period of 2006–2016 is considered the data set is identical with the population of posted TED talks itself.

3.2. Descriptive statistics

This dataset reflects the novelty and strength of the analysis. Before going to empirical analysis, we discuss the descriptive statistics of the data. First, let us look at the main variable of interest (Table 1).

For 2222 talks, the average duration is 13.8 minutes with a standard deviation of 5.9 minutes. The median duration is 14.4 minutes and 90 percent of the talks are completed in at most 20 minutes. There are also a few outliers as well (Figure 1).

The user assigned ratings are simply the adjectives awarded to the talk by the viewers who watched the talk online and rated it. The general website reports the top-two ratings of each talk in hierarchical order. After viewing the talk, a reader is provided with a list of labels to choose from. Surprisingly, only the following 8 among 14 adjectives are awarded by the viewers to talks in top-two place³: Beautiful, Courageous, Fascinating, Funny, Informative, Ingenious, Inspiring and Persuasive.

The duration of the talks differs according to top-two ratings awarded by the viewers, as well as the order of ratings. The descriptive statistics with respect to first rating are as in Table 2:

Table 1. Descriptive statistics of talk duration (in minutes).

Obs.	Mean	Std. Dev.	25th percentile	Median	75th percentile	90th percentile
2222	13.79	5.93	9.47	14.41	17.72	20.02

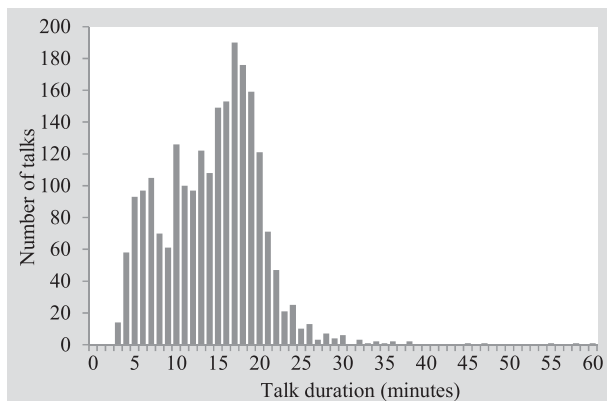


Figure 1. The frequency of talk duration.

Table 2. Descriptive statistics of talk duration by first rating.

Rating:	Mean	Std. Dev.	Freq.
Beautiful	11.60	5.84	138
Courageous	14.35	5.71	67
Fascinating	14.22	6.20	291
Funny	12.00	6.27	156
Informative	13.66	5.91	608
Ingenious	10.22	5.25	105
Inspiring	14.61	5.64	770
Persuasive	16.57	4.39	87
Total	13.79	5.93	2222

We first note the dispersion in terms of the first rating. More than one-third of the talks (770/2222) received ‘inspiring’ as the first rating; while, for about one-quarter of the talks (608/2222), the first rating awarded by the viewers is ‘informative’. We may already see some divergence between the average talk duration and first rating. While the average duration of the talks first-rated as ‘ingenious’ is 10.2 minutes; that of talks first rated as ‘persuasive’ is as high as 16.6 minutes. Considering that 90 percent of the talks have a length of less than or equal to 20 minutes, the difference is sizable.

Similar differences are noticeable also when we consider the second rating (Table 3).

This time, the distribution of the rating is more balanced, still ‘ingenious’ having the least average duration while, talks second-rated as ‘persuasive’ have one of the highest average duration.

3.3. Sequencing of events in users’ behaviour and research questions

The empirical part of the study builds on the following sequence of assumptions:

- Presenter picks a title for the talk,
- Presenter designs the talk with an implicit impact that he/she intends to deliver (or the attitude or message to be delivered), i.e. looking smart, being funny, acting emotionally, being touchy, or pretending to be visionary, etc.,
- Presenter sets an initial talk length while preparing for the talk,

Table 3. Descriptive statistics of talk duration by second rating.

Rating:	Mean	Std. Dev.	Freq.
Beautiful	13.14	7.21	208
Courageous	14.38	5.35	173
Fascinating	13.90	6.03	433
Funny	12.98	7.27	75
Informative	14.92	5.92	444
Ingenious	11.88	5.42	169
Inspiring	13.25	5.67	399
Persuasive	14.03	4.98	321
Total	13.79	5.93	2222

- Presenter interacts with the audience during the talk and the final length of the talk is determined based on the interaction with/feedback from the audience and (perhaps) keeping in mind the message to be delivered or the image/attitude to be communicated,
- The audience leaves the talk with certain feelings and impressions about the context (and the presenter) that might be influenced by the title of the talk, style of the presenter and the duration of the talk,
- Once the talk is posted online, many viewers watch the talk and rate it,
- The ratings awarded by the viewers might be influenced by the style of the presenter, title of the talk as well as by the reaction of the original audience and the duration of the talk.

Within this setting, this study tries to disentangle the relation (if exists) between the title of the talk, the duration of the talk and the ratings awarded by the viewers. Assuming that the simultaneous interaction of the presenter with the original audience and the ex-post interaction of the presenter with the online viewers evolve in a similar manner, then, the results of the seemingly unrelated three branches of the analysis might actually reveal certain regularities regarding the information processing attitudes and online viewing behaviour of individuals considering the attention span and context relationship. The specific research questions that we try to answer are listed as follows:

Question 1: Is there a relationship between talk durations and user ratings?

Question 2: Is there a relationship between talk durations and title keywords?

Question 3: Is there a relationship between user ratings and title keywords?

3.4. Choice of empirical methodology

Having described our data set along with our assumptions about the underlying sequence of events and having outlined our specific research questions, we elaborate in this subsection the empirical framework that has produced the numerical results of the next section. Our choice of methodology is a simple and well-known one in empirical economic and financial research, namely the linear regression approach. Under the assumption of a linear relationship between two continuous variables, this approach allows the researcher to estimate the relationship's intercept and slope parameters. This is done mostly through the

criterion of least squares. Subsequent to estimation, once the differences between the actual values of the dependent variables and the fitted ones (residual or error terms) are distributed normally, the estimated coefficients possess a Student's t distribution and the explanatory power of the fitted relationship can be tested via the ad hoc R^2 or by formally using the F distribution. The linear regression framework is quite versatile owing to its relatively simple theoretical basis and computational tractability.

In the current empirical problem, we seek for the parameters of a relationship between a continuous variable and a categorical one as in duration-rating and duration-keyword, or we seek for the ones of a relationship between two categorical variables as in rating-keyword relation. In both cases we employ a linear regression approach owing to its great flexibility to handle discrete variables as explanatory variables. Such choice of a numerical procedure yields the direction of associations between dependent and independent variables through signs of estimated coefficients as well as providing us with the standard errors of coefficients that are used to conduct significance tests.

Against this background and the structure of the data set, one could resort to an Analysis of Variance (ANOVA) framework so as to F -test the hypothesis that mean duration of talks is the same across ratings (or across keywords) against the alternative hypothesis that at least one rating (or keyword) is associated with a different mean duration. Similarly, while seeking for an association between ratings and keywords, one could suffice with a cross-tabulation along with its overall significance assessment through a chi-square test. In our case, both would be less practical, though. Use of an ANOVA framework, for instance, would not tell us exactly which ratings (or keywords) are associated with longer (or shorter) durations, so requiring a number of sequential testing. The same applies to the case of cross-tabulation. The linear regression approach, being based on the same statistical strand of distributions and tests, allows for a number of tests to be carried out after a single round of estimation.

Eventually, our choice of methodology, we believe turns out to be a good blend of versatility, practicality and communicability across disciplines. In the following, in each research question outlined above, the null hypothesis claims the parameter relevant to the tested relationship to be zero against the alternative of it to be non-zero. So, each test of this form in the following section is concluded through Student's t critical values. In addition, any further tests involving equality versus inequality of parameters are assessed using critical values of F distribution.

4. Results of the empirical analysis

Using the data set and methodology described in the previous section, we devote this section to present our estimates. The following subsections present the relationships between talk durations and ratings, talk durations and keywords in talk titles, and ratings and keywords, in that order.

4.1. Linkages between talk durations and user ratings

The first specification that we consider is a linear regression model where the talk duration is regressed on ratings (first and second rating separately), also controlling for time fixed effects (referring to the month of posting to the Web) (1, 2). As TED talks have a special section called ‘Under 6 minutes’, in order not to bias our results with those observations, we run the regressions on the sample of talks with 6–20 minutes of duration.⁴

$$Duration_{i,t} = \alpha + \sum_{j=1}^8 \beta_j FirstRating_j + \sum_{t=1}^T \delta_t D_t + \varepsilon_{i,t} \quad (1)$$

$$Duration_{i,t} = \alpha + \sum_{j=1}^8 \beta_j SecondRating_j + \sum_{t=1}^T \delta_t D_t + \varepsilon_{i,t} \quad (2)$$

Here, $Duration_{i,t}$ is the duration of talk i aired online in time t ; $FirstRating_j$ are j dummy variables taking the value of 1 if the talk is rated as j th rating, 0 otherwise; D_t are time dummy variables indicating the month of the talk (i.e. first time dummy variable takes the value of 1 if the talk is aired in June 2006 and 0 otherwise).

The estimation results of Specification 1 are presented in Table 4. As there are 8 first-rating variables, one rating becomes the base category in the estimation. In this case, the first rating ‘Beautiful’ is the base. The results indicate that all ratings other than ‘Ingenious’ and ‘Funny’ have

Table 4. Estimation results of specification 1.

First ratings	Dependent var.: Talk Duration
Courageous	2.229*** (0.617)
Fascinating	1.716*** (0.477)
Funny	0.409 (0.568)
Informative	1.432*** (0.436)
Ingenious	−0.948 (0.625)
Inspiring	2.019*** (0.423)
Persuasive	3.814*** (0.555)
Constant	14.48*** (0.828)
Observations	1,748
R-squared	0.169

The first rating ‘Beautiful’ is the base category. The estimation includes time fixed effects (in terms of the month of the talk, output omitted). Robust standard errors in parentheses. *** stand for statistical significance at 1 per cent level.

significantly longer durations, on average, than the talks that are first rated as ‘Beautiful’.

We also test whether the differences between other ratings are also significant. In Table 5, we present the tests for the pairwise equivalence of the coefficients.

Our results robustly indicate that lengths of talks with different first ratings differ in a statistically significant manner for almost all ratings. The distinction is such that talks marked as ingenious are shorter and those marked as persuasive are longer. Quantitatively, there is a difference of 4.8 minutes between these durations (Table 5). Considering that the average talk duration in our sample is 13.8 minutes (Table 1), this suggests that talks that are first rated as persuasive are on average 35% longer than talks that are first rated as ingenious. In a way, a shorter time is needed to sound genius, yet it takes longer to be persuasive. Thus, our findings yield interesting estimates shedding light on people’s attention and appetite regarding an information domain with intellectual ingredients.

As Pizzi, Scarpi, and Marzocchi (2014) argue, when the information is presented in an attribute-based

Table 5. F-tests for the pairwise equivalence of the coefficients of specification 1.

	Beautiful	Courageous	Fascinating	Funny	Informative	Ingenious	Inspiring	Persuasive
Beautiful	–	2.2***	1.7***	0.4	1.4***	−1.0	2.0***	3.8***
Courageous	–	–	−0.5	−1.8***	−0.8	−3.2***	−0.2	1.6***
Fascinating	–	–	–	−1.3***	−0.3	−2.7***	0.3	2.1***
Funny	–	–	–	–	1.0**	−1.4**	1.6***	3.4***
Informative	–	–	–	–	–	−2.4***	0.6***	2.4***
Ingenious	–	–	–	–	–	–	3.0***	4.8***
Inspiring	–	–	–	–	–	–	–	1.8***
Persuasive	–	–	–	–	–	–	–	–

Notes: The values show the difference between the duration of talks: (Column-row), i.e. the talks which were first rated as ‘courageous’ are on average 2.2 minutes longer than talks that are first-rated as ‘beautiful’. ***, ** stand for statistical significance at 1 and 5 percent respectively.

Table 6. Estimation results of specification 2.

Second ratings	Dependent var.: Talk Duration
Courageous	0.868* (0.472)
Fascinating	0.0576 (0.398)
Funny	0.697 (0.613)
Informative	0.828** (0.396)
Ingenious	−0.913* (0.488)
Inspiring	0.487 (0.399)
Persuasive	0.844** (0.415)
Constant	15.49*** (0.904)
Observations	1,748
R-squared	0.137

The second rating 'Beautiful' is the base category. The estimation includes time fixed effects (in terms of the month of the talk, output omitted). Robust standard errors in parentheses. ***, ** and * stand for statistical significance at 1, 5 and 10 percent respectively.

manner, choice of the customers is driven by high-level attributes related to desirability rather than low-level attributes related to feasibility. In our case, we may associate high-level attributes with intellectual/emotional satisfaction expected from the talk (ratings) and low-level attributes with talk duration (as a price for attention devoted). Our results point to some interesting interaction between high-level attributes (ratings) and low-level attributes (duration) of information provided.

Since a typical TED talk is not an academic seminar, nor a thesis defence meeting, nor a business briefing, it can be hard to attribute absolute ratings of genius and persuasive. By the very design of the talks, the presenter might introduce a sharper question at the beginning and puts the punchline fast so as to create a genius impression, and similarly, the presenter might spend more time on philosophical conundrums and provide a concrete explanation at the end to sound persuasive. These are admissible explanations even when the topic of the talk itself does not beg any genius or efforts to persuade. Considering a more trivial alternative, on the other hand, genius talks or persuasive talks may be

genuinely genius or persuasive. Regardless of these alternative views, user-assigned ratings are associated with different average durations.

Second, we turn to the relation between the second-ratings and talk duration. The estimation results of Specification 2 are presented in Table 6.

In Table 7, we present the tests for the pairwise equality of the coefficients after the estimation of Specification 2.

We observe that differences between talk lengths are less with respect to second ratings, which might be indicative of people's ignorance to or degrading of the second ratings. It is also possible that people assign only one rating, i.e. the first one, which is counted thrice when the second and third ratings left null, as indicated by TED organisers. Re-visiting casual human behaviour, such a finding has strong connotations with the motto of 'first impressions do matter'. Equivalently, once a user comes up with a verdict about the first rating, it is possible for her to skip or undermine a second one. This is viable especially when 'the time to assign ratings' is considered as a valuable mental resource. This may also be related to certain attributes of a good or service that customers are consuming. For instance, Brechan (2006) shows that primary attribute regarding the quality of service is more important than the quality of secondary attribute for satisfaction of consumers. Author also suggests that frequent users of service pay more attention to secondary quality attribute than non-frequent users. In that sense, we may argue that TED talk viewers essentially assign the first rating as an indicator of a primary attribute of a talk and only a set of viewers (perhaps more experienced ones) spend more mental resources on choosing a secondary quality attribute for the talk.

Finally, differences between ratings and durations remain under other orderings and specifications (results not presented) in a robust manner. More importantly, our results are intact when talks with durations less than 6 minutes or talks with durations longer than 20 minutes are included to the estimation sample. It is even interesting that significant differences are observed among short talks (under 6 minutes) as well. For instance, the ratings of informative and courageous

Table 7. F-tests for the pairwise equivalence of the coefficients of specification 2.

	Beautiful	Courageous	Fascinating	Funny	Informative	Ingenious	Inspiring	Persuasive
Beautiful	–	0.9**	0.1	0.7	0.8**	−0.9*	0.5	0.8***
Courageous	–	–	−0.8**	−0.2	0.0	−1.8***	−0.4	0.0
Fascinating	–	–	–	0.6	0.8***	−1.0**	0.4	0.8**
Funny	–	–	–	–	0.1	−1.6***	−0.2	0.1
Informative	–	–	–	–	–	−1.7***	−0.3	0.0
Ingenious	–	–	–	–	–	–	1.4***	1.8***
Inspiring	–	–	–	–	–	–	–	0.4
Persuasive	–	–	–	–	–	–	–	–

Notes: The values show the difference between the duration of talks: (Column-row). i.e. the talks which were second rated as 'courageous' are on average 0.9 minutes longer than talks that are second-rated as 'beautiful'. ***, ** and * stand for statistical significance at 1, 5 and 10 percent respectively.

require some additional 25–40 seconds compared to the rating of beautiful.

4.2. A deeper look in the linkages between talk durations and keywords

Subsequent to the first-round estimates of ours above, we focused on wording of talk titles owing to its potential impact on drawing or driving attention. Based on our casual rather than professional experiences, we prepared an unexpectedly rich list of words, phrases or question forms:

- Some are question words found at the beginning of the title: ‘are questions’, ‘can questions’, ‘is questions’, ‘how questions’, ‘how I do something’, ‘how to do something’, ‘what questions’, ‘when questions’, ‘where questions’, ‘which questions’, ‘who questions’, ‘why questions’,
- Some are indicator words: ‘somethings are’ form, ‘somethings are not’ form, ‘something is’ form, ‘something is not’ form, ‘a new something’, ‘a(n) something’,
- Some words indicate presenter’s personal affiliations: ‘statements emphasising I or my’, ‘statements emphasizing we’, ‘let us’,
- Some words refer to the presence of special characters: ‘statements with a numerical figure’, ‘statements in quotes’,
- Others are selected keywords: ‘bad’, ‘big’, ‘brain’, ‘change’, ‘child’, ‘future’, ‘globe’, ‘good’, ‘life’, ‘love’, ‘magic’, ‘math’, ‘music’, ‘my’, ‘myth’, ‘science’, ‘sex’, ‘technology’, ‘time’, ‘world’, ‘you’.

While picking the attention driving words in titles, question forms come first owing to their inherently

challenging sound. Typically, following the initial question of a professor, students start thinking about the question under a controlled level of academic pressure and they tend to be a part of in-class discussions. In the case of being a viewer at a TED talk, however, the viewer sees the title on her screen once and there is no chance that the presenter to direct a question to viewer. Though, an answer to the question posed is granted in the next 18 minutes of watching a talk. Similarly, when the title is signalling about how the presenter does something and how successfully she does it; the element of challenge is embedded therein. Other words we picked are those people are generally interested in, like ‘globe’, ‘change’ or ‘magic’, which are good conversation starters. Hence, a talk with a title including any of the attention driving words, by design or not, might be of a different length compared to talks that do not contain such word in their titles.

Overall, 45 different dummy variables are created identifying each keyword or phrase, the variable taking the value of 1 if the title contains the word and 0 otherwise. The following specification may help test whether the difference in duration for talks containing a keyword and those not containing it is different, where β coefficient measures the difference between the average duration of talks that contain and those that do not contain the specific keyword.

$$Duration_i = \alpha + \beta Keyword_j + \varepsilon_i \quad (3)$$

for $j = 1$ to 45

Table 8 displays the estimated β coefficients for 45 keywords. Having a look at average talk lengths on the basis of inclusion of attention driving words in title we reveal statistically significant differences in talk lengths for 16 out of 45 criteria. The results point to interesting findings. For instance, when ‘how’, ‘when’, ‘which’,

Table 8. Title keywords and talk duration.

Keyword	Coef.	Std. Dev.	Keyword	Coef.	Std. Dev.	Keyword	Coef.	Std. Dev.
are?	0.3	1.4	is not	2.8*	1.5	you	−0.7*	0.4
can?	0.0	1.3	a new	0.0	1.3	big	1.4	1.0
is?	1.3	1.4	a smt	−1.2***	0.4	magic	−2.7***	1.1
how?	0.7*	0.4	I/my ⁺	0.3	0.6	myth	4.3***	0.6
how I?	−0.7	0.8	we	1.3	0.9	math	−0.2	1.2
how to?	−0.3	0.6	let’s	0.5	1.0	science	0.2	0.8
what?	0.4	0.5	numerical	−1.4*	0.7	technology	0.2	1.4
when?	3.1***	1.1	in quotes	−7.4***	0.5	music	−0.6	1.2
where?	4.1***	0.8	world	0.5	0.5	good	0.7	0.8
which?	5.2***	0.1	globe	2.0**	0.9	bad	−0.5	1.0
who?	−1.1	1.8	change	−0.1	0.6	life	−0.1	0.6
why?	1.0***	0.4	child	3.7***	0.5	my ⁺⁺	0.3	0.5
are	1.0	1.1	brain	1.0	0.7	time	0.3	0.9
are not	−6.3***	2.3	future	1.7***	0.6	sex	0.1	1.3
is	0.0	0.6	love	0.6	0.9	other	0.1	0.2

The coefficient shows the difference between the mean duration of talks containing a keyword and the mean duration of talks not containing that keyword (β coefficient in specification 3). ***, ** and * stand for statistical significance at 1, 5 and 10 percent respectively. ⁺Titles starting with ‘I/my’. ⁺⁺Titles containing ‘my’. Here, ‘other’ refers to talks whose title does not contain any of the above.

‘where’ and ‘why’ are considered in turn as leading word in titles, they imply longer talk durations compared to talks that do not contain these starting question words. Similarly, talks containing ‘globe’, ‘child’, ‘future’ and ‘myth’ are in the title are significantly longer than talks that do not contain these words. Meanwhile ‘you’ and ‘magic’ are associated with shorter durations. It is interesting that ‘myth’ and ‘magic’ have opposite impacts on average talk durations.

To make it concrete, talks that contain ‘child’ in the title are on average 3.7 minutes longer than talks that do not contain ‘child’ in the title. Considering the average talk length in the sample, this corresponds to a 27% longer duration. This high difference may not be due to pure coincidence. The title of the talk may actually have an impact on the attention of the audience, whose reactions to the presenter may determine the talk duration.

The revealed statistically significant relation between talk duration and several title words and ratings, leads one to think about two connections: between attention span and context, and between time perception and feelings/emotional space. The attentional states of individuals may change depending on the context and time. For instance, Mark et al. (2014) study the attentional states of workers in information industry by tracking their daily digital activities. They show that the attentional states actually vary with the context of information/task and with the time (day of the week/hour of the day) in a dynamic work setting. In our case, the attitude of the presenter and the title of the talk may alter the attention of the viewers which may in turn affect the duration of the talk performed.

From another perspective, time perception of individuals may also depend on emotional factors. As outlined in the review of Howden (2013) of the book by Hammond (2012), perception of time is altered by seven determinants including attention (concentration), emotions, fear, age, isolation, body temperature and rejection. In a related fashion, Coll-Florit and Gennari (2011) show that it takes longer to process durative events than non-durative events; that the durative events occur in widespread contexts; and that the context-diversity is correlated with processing time, considering online language practising. In this perspective, along with attention, emotional state of the viewers, largely affected by the topic of the talk and the style of the presenter, and the contextual diversity of the talk may affect both the duration of the talk and the ratings awarded to the talk.

4.3. Linkages between ratings and keywords

As a final exercise, we turned to possible linkages between viewer ratings and inclusion of attention driving

words in titles of talks, so we elaborate a slightly angled axis of the central problem in economics of attention: first, we are familiar with the tension between content and context; second, we know that several attributes of information provided online determine audience behaviour; and third, audience in our time is not a passive element of the information architecture; it is an active evaluator, referee or tagger. Then we investigate whether it is possible to find a relationship between user-assigned ratings and wording of talks’ titles in the absence of content, i.e. without the talk itself. We believe, an answer to this question is important for the social scientists in understanding the possible directions for societies in their intellectual journeys.

In this case, we regress the pooled ratings (i.e. beautiful = 1 if first or the second rating is beautiful) on keyword dummy variables as shown in the following specification:

$$Rating_k = \alpha + \beta Keyword_j + \varepsilon_{kj} \quad (4)$$

for $k = 1$ to 8 and $j = 1$ to 45

Our motivation here is to see whether the probability that a talk is rated as ‘such’ differs whether the title of the talk contains a specific word or not. To make it more concrete, one may ask, for instance, if the title of the talk includes ‘music’, does this talk has a higher probability of being rated as ‘beautiful’?

Our empirical results indicate that out of 360 possible relationships (8 ratings by 45 attention driving words) 156 turn out to be statistically significant at the 10 percent level. Indeed, the answer is yes. According to Table 9, if the title of the talk contains ‘music’, the probability of this talk being rated as ‘beautiful’ is 35% higher than talks that do not contain ‘music’ in the title. Also, it is essential to note that Table 9 shows a wide range of starred cells indicating that there is a considerable significant link between the title of the talk and the ratings awarded ex-post.

A case-by-case examination of our estimates reveals interesting linkages. For instance, ‘beautiful’ is positively associated with love, magic, music, life, and my, whereas it is negatively associated with myth and sex, reflecting common aesthetical values. ‘Courageous’ has some positive association with ‘how I’, child and ‘my’. ‘Is something question form’, brain, future, magic and math are positively associated with the rating of ‘fascinating’, where love and magic are positively associated with ‘funny’. ‘Are’, ‘can’, ‘how’, ‘how to’, ‘what’ question forms as well as ‘a new something’, globe, change, brain and future have some positive association with ‘informative’. The negative association between love and ‘informative’ must also be noted. ‘A

something' and 'how I' question forms are positively related to 'ingenious'. The rating of 'inspiring' has a positive relationship with 'how I question form', change, love and life, whereas the relationship reverses for brain and future. Finally, a 'persuasive' rating is positively related to 'why question form', 'let us statement form' and globe, whereas it has a negative relationship with 'how I question form' quite counterintuitively.

We may also read Table 9 through the rows. For instance, if the title starts with 'how' the probability that the talk is rated as 'informative' talk is higher, while the probability that the talk is rated as 'funny' or 'ingenious' is lower, compared to talks whose title does not start with 'how'. Some more examples can be

elaborated as follows: If the title contains the word 'brain', it is more likely to observe a rating of fascinating or informative and less likely to observe beautiful, funny, inspiring or persuasive. Indeed, the difference can be very sizable: the talks titled with brain are 57% more likely to be rated as fascinating than talks that do not contain brain in the title. It is quite likely that those talks aim at unveiling the secrets of human brain and they present some facts not yet known. The audience is less likely to be inspired by a technical talk on brain, or they are not to be persuaded possibly due to a lack of earlier knowledge. A similar observation is valid in the case of 'math' as a title word. It clearly fascinates people, yet they are not likely to see math as something courageous. If the title contains the word 'love', beautiful and funny

Table 9. Attention driving words and ratings (specification 4).

	Beautiful	Courageous	Fascinating	Funny	Informative	Ingenious	Inspiring	Persuasive
are?	-0.07	-0.11***	0.04	-0.01	0.26*	-0.04	-0.25*	0.18
can?	-0.16***	-0.03	-0.01	-0.11***	0.22*	0.02	-0.06	0.12
is?	-0.16***	0.02	0.43***	-0.11***	0.15	-0.01	-0.27*	-0.06
how?	-0.04	-0.03	-0.01	-0.07***	0.22***	-0.09***	-0.05	0.06
how I?	-0.16***	0.26***	-0.23***	0.10	-0.24***	0.14*	0.32***	-0.19***
how to?	-0.1***	0.04	-0.15***	0.04	0.11*	0.00	0.1*	-0.04
what?	-0.04	0.03	0.02	0.00	0.12**	-0.1***	-0.04	0.00
when?	-0.16***	0.56***	0.01	-0.11***	-0.14	-0.13***	0.15	-0.19***
where?	0.09	-0.11***	-0.07	0.14	0.03	-0.13***	0.23	-0.19***
which?	0.84***	-0.11***	-0.32***	-0.11***	-0.47***	-0.13***	0.48***	-0.19***
who?	-0.16***	-0.11***	0.18	0.39	0.03	-0.13***	-0.02	-0.19***
why?	-0.05	0.05	-0.14***	0.01	0.00	-0.14***	0.13**	0.15***
Are	-0.08	0.05	-0.25***	0.13	0.15	-0.05	0.02	0.04
are not	-0.16***	0.39	-0.32***	-0.11***	0.03	-0.13***	-0.02	0.31
Is	0.00	0.03	0.01	-0.02	0.05	-0.05	-0.02	0.00
is not	0.27	0.04	-0.32***	0.04	-0.19	-0.13***	0.34**	-0.04
a new	-0.08	-0.03	-0.09	-0.11***	0.22*	0.02	-0.06	0.12
a smt	0.06*	0.04*	0.01	0.00	-0.11***	0.11***	-0.02	-0.08***
I/my	0.05	0.17***	-0.03	0.03	-0.27***	0.03	0.14**	-0.12***
we	-0.07	-0.02	0.04	-0.11***	-0.02	-0.04	0.03	0.18
let's	-0.1**	0.00	-0.27***	-0.05	0.06	-0.13***	0.11	0.4***
numerical	0.02	-0.11***	-0.11*	0.07	0.10	0.02	0.03	-0.01
in quotes	0.76***	-0.06	-0.33***	0.13	-0.48***	-0.08*	0.25***	-0.19***
world	0.01	-0.02	0.01	-0.02	-0.03	-0.03	0.05	0.03
globe	-0.11***	-0.07	-0.28***	-0.06	0.36***	-0.13***	-0.13	0.43***
change	-0.16***	0.02	-0.23***	-0.01	0.14	-0.07	0.16*	0.14
child	-0.16***	0.29*	-0.22**	-0.11***	0.03	-0.13***	0.28**	0.01
brain	-0.16***	-0.05	0.57***	-0.08***	0.33***	-0.08*	-0.38***	-0.16***
future	-0.14***	-0.09***	0.21***	-0.09***	0.26***	0.00	-0.14*	-0.02
love	0.17*	0.01	-0.12	0.22**	-0.19**	-0.13***	0.16*	-0.11**
you	-0.01	-0.01	-0.01	0.04	0.08**	-0.03	-0.02	-0.05
big	-0.04	-0.05	-0.15	-0.11***	0.3***	-0.01	0.01	0.05
magic	0.18	-0.11***	0.35***	0.29**	-0.42***	0.09	-0.24**	-0.13**
myth	-0.16***	0.29	0.28	-0.11***	-0.07	-0.13***	0.08	-0.19***
math	-0.09	-0.11***	0.39***	0.11	-0.04	-0.06	-0.16	-0.05
science	-0.08	-0.07*	0.06	-0.03	0.03	-0.05	0.02	0.12
technology	-0.01	-0.04	0.18	-0.03	-0.11	0.08	0.05	-0.12*
music	0.35***	-0.06	0.08	0.10	-0.32***	0.12	-0.07	-0.19***
good	-0.05	-0.11***	-0.11	0.05	0.16	-0.13***	0.11	0.08
bad	-0.03	0.02	-0.2*	0.02	0.28*	-0.13***	-0.14	0.19
life	0.12**	-0.02	0.00	-0.01	-0.11*	-0.06	0.15**	-0.08**
my	0.07*	0.18***	0.04	-0.02	-0.19***	0.00	0.09*	-0.18***
time	0.13	-0.01	-0.04	0.05	0.06	-0.1***	-0.08	0.00
sex	-0.16***	0.14	-0.16	0.15	0.11	-0.13***	-0.10	0.15
other	0.00	-0.04***	0.05**	0.00	-0.03	0.05***	-0.05**	0.03

Notes: The rows contain the keywords while the columns contain the ratings. Each cell shows the 'beta' coefficient of the following regression: Rating = $\alpha + \beta \text{Keyword}$. The 'beta' coefficient measures the difference in the probability of a talk being rated as the column when the title of the talk includes the keyword on the row and when not. ***, ** and * stand for statistical significance at 1, 5 and 10 percent respectively. Here, 'other' refers to talks whose title does not contain any of the above.

ratings are more likely to be observed as opposed to persuasive and informative. That is, people might tend to associate 'vivid' ratings with love instead of 'technical', casually cold, ones.

'Why' seems to be a word which transmits the main concern of the talk from the presenter to audience; i.e. the audience assume the responsibility to understand why something happens in a certain way. In the audience ratings fascinating is less likely and persuasive is more likely to encounter.

Probably a more interesting case is that of the word 'child' which straightforwardly is supposed to link with beautiful, fascinating or funny. However, all these ratings are less likely to be observed. Instead, ratings of courageous and inspiring are significantly and positively associated with titles that contain the word 'child'. A deeper look at the titles reveal the background of this finding as talks with 'child' in their titles call for discussions of negative images or struggles to improve children's conditions. Some examples are 'A warrior's cry against child marriage', 'How childhood trauma affects health across a lifetime', 'A child of the state', 'The good news of the decade? We're winning the war against child mortality', 'Teach every child about food', 'The music of a war child' and 'Surprising stats about child car seats'. Obviously, this discussion gives important clues about the context-content relationship.

In a nutshell, one cannot talk about dominantly title-driven outcomes of user-assigned ratings, yet may not totally dismiss such a possibility either. A better understanding of these welcomes further studies.

In addition to duration of the talk, keywords of the title, and the attitude of the presenter; the ratings awarded to the talks may also be influenced by peer effects. Because, when the viewer selects the talk that she wants to view, she already knows the first two ratings of the talk in advance. Even if she does not attend to this available information, she still has a chance to view the ratings awarded after watching the video of the talk and before casting her own feelings about the talk and the presenter. In a recent study that analyses the possible impact of social information on the choices of individuals in an experimental setting, Delfino, Marengo, and Ploner (2016) show that the choices of individuals are affected by information regarding the choices of peers. Such imitation behaviour is stronger when the social information corresponds to the sample average choice. This evidence is provided from a setting where the choice is related to an investment and individuals expect to gain monetary returns. Although such a peer effect may be considered as a caveat, we think that the probability of an imitative behaviour emerging in the current setting of rating TED talks – an endeavour that neither incurs

a monetary cost nor promises a monetary return – is very limited.

Finally, the combination of 8 ratings with 45 different indicators may seem to produce an excessive number of scenarios. Given that not all 360 combinations produce statistically significant results, such concerns may pile up further. However, behind our extensive coverage of different combinations lies the virtue of big data research, to which authors truly believe in. Essentially, not all the collected information will be useful, but some information may actually be quite useful, which may not be revealed if such efforts on looking at a large set of events are not spent. Nonetheless, caution must be exercised in order to avoid the potential negative sides of a large amount of information in terms of accuracy and depth of insight. In our case, the advantages of large data use weigh more as the data is naturally produced by speakers and viewers consciously, rather than by a machine registry of any sort.

5. Discussion and concluding remarks

Herbert Simon's words putting forward a decline in attention following a wealth of information could not be better translated into human experience than by the Internet did in the last decades. Traditionally higher money costs of acquiring information were replaced by the amount of attention paid by people, once the monetary costs were virtually zeroed. In today's world, it does not seem wrong to claim that attention is a scarce resource and how it is allocated to an enormous count of information sources deserves scholar attention.

In this paper, we use a naturalistic data environment to analyse the information selecting and processing attitudes of individuals in the context of online viewing behaviour. By using the titles, duration and viewer-assigned ratings of more than two-thousand TED talks, we analyse whether the length of an information stream (talk duration) is related to some user-assigned attributes (ratings); whether there is a linkage between the length of information streams (talk duration) and some specific attention driving factors (words appearing in talks' titles); and whether the ex-post user-assigned ratings and ex-ante wording of talks' titles are connected. Use of the domain of TED talks recently became popular in several disciplines and a fruitful venue for many researchers, mainly owing to its good design, wide coverage, core content and rich provision of metadata. Employing those, we document statistically significant interactions on each of the three questions considered.

Our results reveal certain regularities regarding information processing attitudes and online viewing behaviour retrieval attitudes of individuals, the attention and

context interaction and the link between time perception and emotional state. The revealed regularities can be consolidated to following items so as to put them in perspective:

- [1] Regardless of the type of user rating, i.e. the first versus second, talk durations statistically significantly differ with respect to ratings, as in the case of shorter time to be needed to sound genius yet it takes longer to be persuasive. However, the differences between talk lengths are less with respect to second ratings as documented in the first subsection of section 4.

This line of findings is indicative of, at a conscious or subconscious level, viewers' associations of their value assessments to viewing time spent. More importantly, their first impressions (first ratings) are associated with talk durations more as compared to secondary assessments (second ratings). That is, the cost of viewing (time spent on videos) in relation to benefit derived (information, message or joy from videos) seems to be an important determinant of their primary assessments (first ratings). By the sequencing of events here (view first rate later) associations between talk durations and ratings seem to reflect some causality.

- [2] Talk lengths do have statistically significant associations with attention driving words (keywords) in talk titles as documented in the second subsection of section 4.

The estimates of section 4 that have linked talk lengths to attention driving words seem to have uncovered an important dimension of the problem at hand. The simple reason for this is the absence of online viewers in determining both the lengths and titles of talks. Simply speaking, TED talks presenters (root providers of information here) seem to have a delivery or marketing strategy, which is not surprising but measurable even in the absence of a full knowledge of more than two thousand videos.

- [3] User ratings have some statistically significant linkages with attention driving wording in titles as documented in the third subsection of section 4.

Once they have viewed the talks with the knowledge of titles at hand, the viewers assign ratings to talks which have some statistically significant linkage to titles' attention driving wording. This accounts for 156 combinations among the 360 combinations of 8 ratings and 45 attention driving words. Equivalently, higher than 40

percent of the time we are able to measure a significant association.

All in all, the cost of handling information (talk duration), how the information has been packaged (attention driving words) and users' assessments (first and second ratings) yield a rich set of statistically supported measurements, and they do so in the absence of direct visual assessments of the videos or their transcripts.

The current study points at a challenging future research agenda in which researchers maintain a content-orientation by including the audio-visual-textual substance of talks. With regard to that, a thorough assessment of videos including the changing pace and other emotional markers of presenters, occurrences of jokes within talks, visual projections used in talks, use of specific words or phrases in talks and information on the co-appearance of talks in the same TED talks session would be enlightening. Such studies might require, though, a more ambitious collection of Internet metadata especially with regard to completion of talks by online viewers and they could find out trigger words that cause termination of viewing, or they could reveal the effects of on-site viewers' stance (as appeared in videos) on online viewing behaviour. Such a study would require a much more complex data collection and analysis, which is not the scope of the current study. Still, what is amazingly documented here is the ability of sole metadata to provide us with strong findings about how human beings handle information in this new age of knowing things, which certainly feeds the appetite of researchers to delve deeper more into context analysis of talks as well.

Notes

1. The talk was given at a local TEDx event, produced independently of the TED conferences, as disclaimed by the disseminators.
2. We collect all the information available on the general pages where a pile of talks is located. Although more information is available in the private pages of each talk, we prefer not to engage in such an endeavor of collecting more detailed data in this study.
3. The full list of available adjectives includes Courageous, Inspiring, Jaw-dropping, Informative, Beautiful, Persuasive, Fascinating, Confusing, Obnoxious, Ingenious, OK, Longwinded, Unconvincing, Funny.
4. With this approach, we still cover the great bulk of the talks in our sample (we omit 10% of the talks that are >20 minutes; and 12% of the talks that are ≤6 minutes).

Disclosure statement

No potential conflict of interest was reported by the authors. The views expressed in the paper belong to the authors and

they do not necessarily reflect those of the Central Bank of the Republic of Turkey, or its staff.

References

- Brechan, I. 2006. "The Different Effect of Primary and Secondary Product Attributes on Customer Satisfaction." *Journal of Economic Psychology* 27 (3): 441–458.
- Broadbent, D. 1958. *Perception and Communication*. London: Pergamon Press.
- Cettolo, M., C. Girardi, and M. Federico. 2012. "WIT3: Web Inventory of Transcribed and Translated Talks." In *Proceedings of the 16th EAMT Conference*, 28–30 May 2012, Trento, Italy.
- Coll-Florit, M., and S. P. Gennari. 2011. "Time in Language: Event Duration in Language Comprehension." *Cognitive Psychology* 62 (1): 41–79.
- Delfino, A., L. Marengo, and M. Ploner. 2016. "I did it Your way. An Experimental Investigation of Peer Effects in Investment Choices." *Journal of Economic Psychology* 54: 113–123.
- Deutsch, J., and D. Deutsch. 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.
- Fiksdal, A. S., A. Kumbamu, A. S. Jadhav, C. Cocos, L. A. Nelsen, J. Pathak, and J. B. McCormick. 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., and X. Bai. 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.
- Hammond, C. 2012. *Time Warped: Unlocking the Mysteries of Time Perception*. Edinburgh and London: Canongate. ISBN 978-1-84767-790-7.
- Howden, D. 2013. "Review of *Time Warped: Unlocking the Mysteries of Time Perception*, by Claudia Hammond." *Journal of Economic Psychology* 34: 304–305.
- Hsieh, T. L., H. H. Lo, H. H. Hu, and C. C. Chang. 2015. "The Effect of Information Design on Cognitive Processing of Website Navigation." *Human Factors and Ergonomics in Manufacturing & Service Industries* 25 (5): 548–558. doi:10.1002/hfm.20568.
- Huberman, B. A. 2009. "Social Attention in the Age of the Web." *Working Together or Apart: Promoting the Next Generation of Digital Scholarship* 62: 62–69.
- Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., and A. Tversky. 1979. "Prospect Theory: an Analysis of Decision Under Risk." *Econometrica* 47 (2): 263–292.
- Lanham, R. A. 2006. *The Economics of Attention: Style and Substance in the Age of Information*. Chicago: University of Chicago Press.
- Lee, L. H., Y. C. Juan, W. L. Tseng, H. H. Chen, and Y. H. Tseng. 2015. "Mining Browsing Behaviors for Objectionable Content Filtering." *Journal of the Association for Information Science and Technology* 66 (5): 930–942.
- Lopes, J., I. Trancoso, and A. Abad. 2011. "A Nativeness Classifier for TED Talks." In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 22–27 May 2011, Date Added to IEEE Xplore: 12 July 2011, doi:10.1109/ICASSP.2011.5947647.
- Mark, G., S. T. Iqbal, M. Czerwinski, and P. Johns. 2014. "Bored Mondays and Focused Afternoons: The Rhythm of Attention and Online Activity in the Workplace." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April, 3025–3034. ACM.
- Miller, G. A., E. Galanter, and K. H. Pribram. 1960. *Plans and the Structure of Behaviour*. New York: Holt, Rinehart & Winston.
- 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. doi:10.1108/OIR-05-2015-0139.
- Pappas, N., and A. Popescu-Belis. 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, doi:10.1145/2484028.2484116.
- Pizzi, G., D. Scarpi, and G. L. Marzocchi. 2014. "Showing a Tree to Sell the Forest: The Impact of Attribute- and Alternative-Based Information Presentation on Consumers' Choices." *Journal of Economic Psychology* 42: 41–51.
- Romanelli, F., J. Cain, and P. J. McNamara. 2014. "Should TED Talks Be Teaching Us Something?." *American Journal of Pharmaceutical Education* 78 (6), Article 113.
- Rousseau, A., P. Deleglise, and Y. Esteve. 2014. "Enhancing the TED-LIUM Corpus with Selected Data for Language Modeling and More TED Talks." In *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.
- Segev, E., and N. Ahituv. 2010. "Popular Searches in Google and Yahoo!: A 'Digital Divide' in Information Uses?" *The Information Society* 26 (1): 17–37.
- Simola, J., J. Hyönä, and J. Kuisma. 2014. "Perception of Visual Advertising in Different Media: From Attention to Distraction, Persuasion, Preference and Memory." *Frontiers in Psychology* 5 (1208), paper 5. doi:10.3389/fpsyg.2014.01208.
- Simon, H. A. 1971. "Designing Organizations for an Information-Rich World." In *Computers, Communication, and the Public Interest*, edited by M. Greenberger, 38–72. Baltimore, MD: The Johns Hopkins Press.
- Stephen, W. 2015. "How to Sound Smart in Your TEDx Talk?." talk delivered in TEDxNewYork and aired via YouTube on January 15, 2015, last accessed online on January 28, 2017 at <https://www.youtube.com/watch?v=8S0FDjFBj8o>.
- Sugimoto, C. R., M. Thelwall, V. Larivière, A. Tsou, P. Mongeon, et al. 2013. "Scientists Popularizing Science: Characteristics and Impact of TED Talk Presenters." *PLoS ONE* 8 (4): e62403. doi:10.1371/journal.pone.0062403.

- Treisman, A. 1964. "Selective Attention in man." *British Medical Bulletin* 20 (1): 12–16.
- Tsou, A., M. Thelwall, P. Mongeon, and C. R. Sugimoto. 2014. "A Community of Curious Souls: An Analysis of Commenting Behavior on TED Talks Videos." *PLoS ONE* 9 (4): e93609. doi:10.1371/journal.pone.0093609.
- Wook, T. S. M. T., and S. S. Salim. 2014. "Visual Interface for Searching and Browsing Children's WebOPAC." *Universal Access in the Information Society* 13 (4): 367–385.
- Zhang, L., T. Q. Peng, Y. P. Zhang, X. H. Wang, and J. J. Zhu. 2014. "Content or Context: Which Matters More in Information Processing on Microblogging Sites." *Computers in Human Behavior* 31: 242–249.