Kerkhof, Anna

Conference Paper

Advertising and Content Differentiation: Evidence from YouTube

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2019: 30 Jahre Mauerfall - Demokratie und Marktwirtschaft - Session: Empirical Industrial Organisation I, No. C19-V1

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Advertising and Content Differentiation: Evidence from YouTube*

Anna Kerkhof†

Job Market Paper

This version: September 4, 2019

Abstract

Does advertising revenue increase or diminish content differentiation in media markets? This paper shows that an increase in the technically feasible number of ad breaks per video leads to an increase in content differentiation between several thousand YouTube channels. I exploit two institutional features of YouTube's monetization policy to identify the causal effect of advertising on the YouTubers' content choice. The analysis of around one million YouTube videos shows that advertising leads to a twenty percentage point reduction in the YouTubers' probability to duplicate popular content, i.e., content in high demand by the audience. I also provide evidence of the economic mechanism behind the result: popular content is covered by many competing YouTubers; hence, viewers who perceive advertising as a nuisance could easily switch to a competitor if a YouTuber increased her number of ad breaks per video. This is less likely, however, when the YouTuber differentiates her content from her competitors.

JEL Codes: D22, L15, L82, L86

Keywords: advertising, content differentiation, economics of digitization, horizontal product differentiation, long tail, media diversity, user-generated content, YouTube

*I thank Johannes Muenster for helpful comments, discussion, and guidance. Moreover, I thank Maja Adena, Felix Bierbrauer, Joerg Claussen, Milena Djourelova, Ruben Durante, Matthew Ellman, Ruben Enikolopov, Thomas Fackler, Marcel Garz, Fabrizio Germano, Julian Harke, Matthias Heinz, Ulrich Kaiser, Reinhold Kesler, Tobias Kretschmer, Ulrich Laitenberger, Jonas Loebbing, Leonardo Madio, Sebastian Mertesacker, Axel Ockenfels, Martin Peitz, Christian Peukert, Susanne Prantl, Sarah Recht, Maximilian Schaefer, Robert Somogyi, Frederic Thenee, Kevin Tran, Dainis Zegners, and participants of the 3rd Workshop on the Economics of Media Bias, the Media Economics Reading Group in Barcelona UPF, the Applied Microeconomics Seminar of the University of Cologne, the WZB Seminar, the Digital Economics Workshop Lisbon, the ICT Paris, the SMYE Brussels, the Digital IO Workshop Louvain-la-Neuve, the ORG Seminar Munich, the Barcelona Summer Forum, the DICE Research Seminar, the ICT Mannheim, the EEA Manchester, and the EARIE Barcelona for helpful comments. All errors are my own.

†University of Cologne. E-mail: anna.kerkhof[at]wiso.uni-koeln.de.

1. Introduction

Modern societies pursue media diversity and content differentiation between media outlets for at least two reasons. First, a diverse media landscape sustains democracy, freedom, and the public discourse (Downs, 1957; Coase, 1974; Gentzkow et al., 2016; Puglisi and Snyder, 2016). Second, preferences over media content differ substantially across different groups of consumers: men and women prefer different types of media content, as do young and old people, and consumers with different socio-economic backgrounds – the more differentiated the content in media markets, the more likely it is that all consumers' preferences are served (Waldfogel, 2007). Policy makers, too, appreciate the value of content differentiation and undertake great efforts to achieve and maintain diverse media markets.[1]

The media outlets' typical business model – generating advertising revenue instead of charging their consumers a monetary price – provokes a persistent debate about the consequences of advertising on content differentiation, however. Media scholars and the German supreme court argue that advertising revenue induces media outlets to duplicate popular content – i.e., content in high demand by the audience – to sell a maximum number of eyeballs to advertisers (e.g., Herman and McChesney, 1997; Hamilton, 2004; McChesney, 2004; Bundesverfassungsgericht, 2007). Predictions from economic theory, on the other hand, are ambiguous. Pioneering models on media outlets' content choice by Steiner (1952), Beebe (1977), and Gabszewicz et al. (2001) follow the above argumentation and predict that advertising leads to minimum differentiation à la Hotelling (1929). More recent models, however, acknowledge that many consumers perceive advertising as nuisance and thereby as a "price" they have to pay (Wilbur, 2008; Huang et al., 2018; Anderson and Jullien, 2016). Taking this into account leads to the opposite prediction: when incentivized by ad revenue, media outlets prefer to differentiate their content from each other to soften competition in the ad "price." Does advertising increase or diminish content differentiation in media markets? Empirical evidence on this question is scarce.

This paper studies the effect of advertising on content differentiation on YouTube – the world's most visited user-generated content platform[2] – to resolve the open question. I exploit two features of YouTube's monetization policy to identify the causal effect of advertising on the YouTubers' probability to duplicate popular content, where I define popular content as the content that attracts the largest number of views. First, I make use of the "ten minutes trick", which is a discontinuity in YouTube's mapping from video duration to the technically feasible number of ad breaks per video. If a video is shorter than ten minutes, YouTubers can permit for exactly one ad break in it. If the video is ten minutes or longer, YouTubers face no such limitation. Second, the ten minutes trick was hidden until Oct 2015, when YouTube launched a new ad break tool that made its existence prominent.

[1]The European Council, for instance, has recently passed official guidelines for the protection of media diversity in the EU (CM/Rec(2018)1).

[2]See `www.alexa.com/topsites` (July 2019)

To identify the effect of advertising on the YouTubers' content choice, I consider only YouTubers who produced short videos before Oct 2015, since they were likely to be unaware of the ten minutes trick before the new ad break tool was launched. Based on this sample, I classify a YouTuber as "treated" if she could increase her feasible number of ad breaks by increasing her share of videos that are ten minutes or longer *after* Oct 2015, and as untreated otherwise. Then, I compare the change in the probability to duplicate popular content before and after Oct 2015 for YouTubers in the treatment relative to the control group in a difference-in-differences framework.

Since the YouTubers have perfect control over their videos' duration, they might self-select into the treatment group. To account for endogeneity in the YouTubers' treatment status, I use their *median video duration before Oct 2015* – i.e., their "closeness" to the ten minutes threshold before Oct 2015 – as an instrument for being treated. The YouTubers in the sample did not choose their videos' duration *before* Oct 2015 bearing the ten minutes trick in mind, because they were unaware of the feature. As a result, a YouTuber's median video duration before Oct 2015 is uncorrelated to omitted variables that drive self-selection into the treatment group (e.g., commercial interests). The YouTubers' median video duration before Oct 2015 is furthermore uncorrelated to the popularity of her videos' content. On the other hand, extending their videos' duration to ten minutes or more is easier for YouTubers who were "closer" to the threshold *before* Oct 2015, i.e., median video duration before Oct 2015 is correlated to the YouTubers' (potentially endogenous) treatment status. A broad range of validity checks supports the identification strategy.

The analysis of around one million YouTube videos shows that an increase in the feasible number of ad breaks per video leads to a twenty percentage point reduction in the YouTubers' probability to duplicate popular content. The effect size is considerable: it corresponds to around 40% of a standard deviation in the dependent variable and to around 50% of its baseline value. The large sample size allows me to conduct several subgroup analyses to study effect heterogeneity. I find that the positive effect of advertising on content differentiation is driven by the YouTubers who have at least $1,000$ subscribers, i.e., the YouTubers whose additional ad revenue is likely to exceed the costs from adapting their videos' content. In addition, I find heterogeneity along video categories: some categories are more flexible in terms of their typical video duration than others, hence, exploiting the ten minutes trick is more easy (e.g., a music clip is typically between three and five minutes long and cannot be easily extended). A battery of robustness checks confirms these results.

Recent economic models on content choice in media markets acknowledge that consumers perceive advertising as a nuisance and similar to a "price" they have to pay; media outlets differentiate from each other to avoid ruinous competition in the ad "price" as a consequence (see Anderson and Jullien, 2016, for a survey). In the second part of the paper, I show that the avoidance of ad "price" competition is a plausible economic mechanism behind my main results. First, I demonstrate that popular content – i.e., content in high demand – is also *supplied* by many YouTubers. Thus, viewers could easily switch to a competitor if a YouTuber increased her ad "price." Switching becomes less likely, however, when the YouTuber uploads content that is less popular and thereby covered by fewer competitors. Next, using the empirical framework from above, I show that an increase in the feasible number of ad breaks leads to a twenty percentage point reduction in the YouTubers' probability to upload content that is covered by many other YouTubers, too. Finally, I support this result by demonstrating that the audience of YouTubers who could increase the feasible number of ad breaks per video becomes more stable, i.e., the viewers become less likely to switch to competitors. I find no evidence for other economic mechanisms behind my results.

The paper contributes to two recent debates. First, I advance the discussion on the effect of advertising on content differentiation in media markets. To my knowledge, this is the first paper that provides evidence of a causal *positive* effect of advertising on content differentiation, whereby it challenges the widespread opinion that the media inefficiently duplicate popular content when incentivized by ad revenue. This is a major insight, especially because the media's options to generate ad revenue are often subject to external regulation.[3]

Second, my results contribute to recent discussions about the effect of digitization on content differentiation and diversity in media markets (Waldfogel, 2017, 2018). The traditional cost structure of media markets – fixed costs are high and marginal costs are low – impedes media diversity, because the number of outlets that can co-exist is limited. Goldfarb and Tucker (2019), however, point out that digital technology has "reduced the cost of storage, computation, and transmission of data" (p.3). As a result, online media outlets can afford to provide niche content, while enhanced search technologies enable consumers to find it – a phenomenon that Anderson (2006) summarizes as "the long tail."[4] YouTube serves as a point in case to study the determinants of content differentiation in digital media markets in general when fixed costs are low (e.g., online news markets or alternative user-generated content platforms). In particular, technology alone may not ensure a more diverse media landscape: although large number of media outlets can co-exist, they might duplicate the most popular content, while niche preferences remain unserved. My paper shows that advertising provides additional incentives for media outlets to differentiate their content that – when falling on the fertile ground of digitization – can help to increase media diversity.

I contribute to two additional strands of literature. First, my paper adds to the extensive work on horizontal product differentiation (see, e.g., Graitson, 1982; Gabszewicz and Thisse, 1986; Lancaster, 1990; Anderson et al., 1992), which shows that firms' degree of product differentiation is determined by two contrasting effects. On the one hand, a direct effect induces firms to move closer to their competitors to increase their consumer base, leading to *minimum differentiation* (Hotelling, 1929). On the other hand, a strate-

[3]The Audiovisual Media Services Directive, for instance, requires that the proportion of television advertising and teleshopping spots within a given clock hour shall not exceed 20% (Article 23 §1).

[4]See also Brynjolfsson et al. (2003, 2011) for a discussion on the long tail and how consumer surplus benefits from increased product variety.

gic effect prompts firms to move away from their competitors to soften price competition, which leads to *maximum differentiation* (d'Aspremont et al., 1979; Economides, 1986).[5] Accordingly, models on content differentiation in media markets that ignore consumers' ad aversion find that advertising leads to minimum content differentiation (Steiner, 1952; Beebe, 1977; Gabszewicz et al., 2001, 2002; Garcia Pires, 2014; Behringer and Filistrucchi, 2015). Models that acknowledge the conceptual equivalence between direct prices and consumers' nuisance costs from advertising, in contrast, predict that media outlets prefer to differentiate from each other to avoid ruinous competition in the ad "price" (Bourreau, 2003; Dukes, 2004; Gabszewicz et al., 2004; Peitz and Valletti, 2008; Anderson and Jullien, 2016).[6]

My paper provides causal empirical evidence for the theoretical considerations from this literature. While a related paper by Seamans and Zhu (2014) shows that an increase in subscription prices is correlated to a higher degree of content differentiation, I demonstrate that an increase in the feasible number of ad breaks per video leads to content differentiation, because YouTubers want to soften competition in the ad "price." Most closely related to my work is Sun and Zhu (2013), who study the introduction of an ad-revenue-sharing program on a major Chinese online platform and find that advertising leads to the duplication of popular content. Our results do not contradict each other, though. While ad breaks before or during YouTube videos are a true nuisance to viewers, Sun and Zhu (2013) explicitly state that the ads appearing on the bloggers' posts are not intrusive (p. 2317), which means that only a direct, but no strategic effect operates in their setting. The papers can therefore be seen as complements supporting the plausibility of each other's results.

In addition, my work makes three contributions to the literature on user-generated content (see Luca, 2016b, for a survey). First, I present a novel empirical strategy to identify causal effects on a user-generated content platform. While existing approaches use variation in institutional features *across* platforms (e.g., Chevalier and Mayzlin, 2006; Mayzlin et al., 2014), *within* platforms (Anderson and Magruder, 2012; Luca, 2016a), or conduct randomized experiments (Bond et al., 2012; Aral and Walker, 2012), I exploit two distinctive features of YouTube's monetization policy to identify the causal effect of advertising on the YouTubers' content choice. Second, I apply this novel identification strategy to a unique dataset of newly collected data on several thousand German YouTubers with more than a million videos that have not been investigated before. Third, my paper explores how monetization affects user-generated content. Since many other user-generated content platforms such as Wikipedia, TripAdvisor or Twitter do not allow their contributors to earn money, YouTube offers a unique environment to study this question. Previous

[5] De Palma et al. (1985) show that if the consumers are sufficiently heterogeneous in terms of their taste parameter, the direct effect prevails.

[6] Gal-Or and Dukes (2003) find minimal differentiation even if consumers are ad averse, but the result is driven by the assumption of informative advertising. When the outlets minimally differentiate their content, advertisers choose lower levels of advertising, because the consumers are ad averse. This implies lower levels of product information to consumers, whereby the advertisers gain higher margins on their products. As a result, the media outlets can set higher prices for advertisers (p.292).

analyses show that users contribute to user-generated content platforms for two main reasons: reputation (Wang, 2010; Anderson et al., 2013; Easley and Ghosh, 2013) and beliefs about a high impact of their contributions (Zhang and Zhu, 2011). These motives are non-pecuniary and render it unclear whether the YouTubers react to economic incentives at all. My results demonstrate that economic considerations matter. In particular, when incentivized by ad revenue, the YouTubers are willing to deviate from the content they provided before. Moreover, I show that ad revenue does not necessarily improve the YouTubers' video quality. Although the number of views goes up when a video has more ad breaks, the relative number of likes decreases.

The remainder of the paper is organized as follows. Section 2 provides background information on YouTube, its monetization policy, and the institutional features that the empirical strategy builds on. A stylized example introduces the central ideas of identification in Section 3, before I illustrate the data collection process and how I construct a dataset that is suitable for the analysis in Section 4. Section 5 discusses the details of the empirical strategy; the results are presented in Section 6. Next, in Section 7, I explore the economic mechanism that drives these results. Section 8 studies content differentiation in the aggregate; Section 9 investigates changes in video quality. Section 10 concludes.

2. YouTube: Background

2.1. Platform, audience, and contributors

YouTube is a video sharing platform founded in 2005 and acquired by Google in 2006. Its reach is tremendous: with 800 million unique users and 15 billion visits per month, it is the second-most popular website in the world (after google.com).[7] As of Oct 2018, several billion hours of video content from YouTube are watched every day.[8]

YouTube is based on user-generated content. While unregistered users are limited to watching, registered users can upload, share, and comment on videos. Registered users who upload videos on a regular basis are called *YouTubers*; YouTubers, in turn, operate a YouTube *channel* under their user name to distribute their videos.[9]

2.2. Monetization

YouTubers have the option to monetize their content; in particular, they can generate advertising revenue by permitting YouTube to show ads to viewers before or during their videos. However, while YouTubers can permit that ads *may* be shown, YouTube's algorithm determines *if* and *which* ad is displayed to a particular viewer. Thus, there is no direct relationship between YouTubers and advertisers.[10] According to anecdotal evidence

[7]See `www.alexa.com/siteinfo/youtube.com` (Oct 2018).

[8]See `www.youtube.com/yt/about/press/` (Oct 2018).

[9]I use the terms "YouTuber" and "channel" synonymously; cases where one YouTuber operates several channels are rare.

[10]In addition to permitting for ad breaks in their videos, YouTubers might also earn money through product placement and affiliate links. In this case, there exists a contractual basis with the advertiser.

– official statistics do not exist – YouTubers earn about three to five USD per 1,000 views per ad per video.[11]

Monetization via ad breaks is not open to all YouTubers, though. First, a YouTuber's content must be advertiser-friendly, i.e., free of violence, sex, and crime.[12] In early 2017, YouTube introduced a new policy of automated demonetization of non-advertiser-friendly content (also known as "adpocalypse") that aims at videos on sensitive social issues, tragedy, or conflict; many YouTubers reported losing more than half of their income as a result.[13] Second, while not bounded to a subscriber threshold before, YouTube disabled the monetization option for YouTubers with fewer than 1,000 subscribers in Feb 2018. This policy, too, is a reaction to advertisers' complaints about their products appearing next to dubious video content.[14] The subscriber threshold, YouTube argues, gives them enough information to determine the validity of a YouTuber's channel and to confirm that it is following the YouTube community guidelines and advertiser policies.[15]

2.3. The ten minutes trick

YouTube's monetization policy exhibits one distinctive feature, which is known as the "ten minutes trick." The ten minutes trick refers to a discontinuity in YouTube's mapping from a video's duration to the technically feasible number of ad breaks that the YouTuber can permit. If a video is shorter than ten minutes, YouTubers can permit for exactly one ad break in it. If, on the other hand, the video is ten minutes or longer, YouTubers face no technical restriction on the number of ad breaks.[16] Hence, the ten minutes trick can be summarized as

$$feasible\, number\, of\, ad\, breaks = \begin{cases} 1 & if\, video\, duration < 10\, min \\ \infty & if\, video\, duration \geq 10\, min. \end{cases} \tag{1}$$

While the ten minutes trick had long been a hidden feature, it gained sudden prominence in Oct 2015, when YouTube launched a new ad break tool for YouTubers.[17] The tool had two effects. First and foremost, it made the ten minutes trick apparent. In its old version, only a small additional input box would appear for videos exhibiting the ten minutes threshold (A in Figure 1). In contrast to that, the option to embed additional ad breaks is now permanently visible and points YouTubers to its existence (B in Figure 2). Second, editing additional ad breaks became less cumbersome. The new tool allows YouTubers to drag ad breaks back and forth on their video time line and it also offers a preview option to check whether an ad appears at an appropriate point in time during the video (C and

[11]See `influencermarketinghub.com/how-much-do-youtubers-make/` (Dec 2018).
[12]See `support.google.com/youtube/answer/6162278?hl=en` (Dec 2018).
[13]See `nymag.com/intelligencer/2017/12/can-youtube-survive-the-adpocalypse.html` (Dec 2018).
[14]See `turbofuture.com/internet/YouTube-Screwed-Small-YouTube-Channels-With-Their-New-Memorization-Policy` (Dec 2018).
[15]`support.google.com/youtube/answer/72857?hl=en\&ref_topic=6029709` (Dec 2018).
[16]`support.google.com/youtube/answer/6175006?hl=en` (Oct 2018).
[17]See `www.youtube.com/watch?v=z58Ed6q6xQg` (Oct 2018).

D in Figure 2). The old version, in contrast, required typing and re-typing the point in time where the ad breaks were supposed to appear (A in Figure 1).

3. Identification: Stylized example

An ideal experiment would randomly assign some YouTubers to the option of showing just one, and others to the option of showing several ads per video to their viewers, and then compare the groups' probabilities to upload popular content. Given that the YouTubers' real life monetization settings are endogenous, however, the identification of a causal link from advertising to content choice requires a thoughtful empirical strategy. Though highly stylized, this section illustrates how combining the ten minutes trick with the launch of the new ad break tool yields variation in the YouTubers' feasible number of ad breaks per video that I exploit to identify the causal effect of interest.

Figure 3 illustrates YouTube's mapping from video duration to the technically feasible number of ad breaks per video as described in Section 2. Consider three hypothetical YouTubers A, B, and C *before* Oct 2015, where A's videos are very short, B's videos are close to but still below the ten minutes threshold, and C's videos are longer than that. Hence, while A and B may only permit for one ad break per video, C faces no such limitation. Note that this is no regression discontinuity setting, because the YouTubers have perfect control over their videos' duration. In particular, C could have chosen her videos' duration strategically to benefit from the jump in the feasible number of ad breaks per video.

Next, consider the launch of the new ad break tool in Oct 2015. While C is unaffected, A and B realize that they can increase the feasible number of ad breaks per video by uploading videos that are ten minutes or longer. Pushing her video duration beyond the threshold, however, is easier to accomplish for B than for A. The key identifying assumption is that although a YouTuber has perfect control over her videos' duration, A and B, who were initially ignorant of the threshold's existence, did not choose their videos' *distance* to the ten minutes threshold having the discontinuity in mind. As a consequence, the cost of moving beyond the threshold *after* it became prominent – and thereby also the probability to actually do so – is orthogonal to unobserved characteristics such as, for instance, commercial incentives that may also drive a YouTuber's decision to increase her feasible number of ad breaks.[18]

I exploit the variation in the YouTubers' cost to move beyond the threshold as follows. First, I consider only YouTubers like A and B, i.e., YouTubers who were "left" to the ten minutes threshold before Oct 2015. Then, I compare the change in the probability

[18] To be precise, A and B could correspond to three types of YouTubers: (i) those who did not know about the threshold, as discussed above, (ii) those who knew about the threshold, but found it too cumbersome to permit for additional ad breaks, and (iii) those who knew but did not want to increase their videos' duration. The logic that applies to YouTubers in group (i) holds for YouTubers in group (ii) as well. YouTubers in group (iii) can be interpreted as "never takers", see Section 5.2.2 for a discussion of IV heterogeneity.

to upload popular content before and after Oct 2015 of YouTubers who could increase the feasible number of ad breaks per video by uploading videos that are ten minutes or longer (treatment group) to YouTubers who did not do so (control group) in a difference-in-differences framework. Finally, I account for self-selection into the treatment group by using a YouTuber's "closeness" to the ten minutes threshold before Oct 2015 as an instrument for her treatment status. Thus, my empirical strategy boils down to exploit variation YouTubers who were close to the threshold before Oct 2015 to YouTubers who were further away from it (in contrast to comparing YouTubers just left to the threshold to YouTubers just right to it, as one would do in a regression discontinuity design). A detailed discussion of the empirical strategy follows in Section 5.

4. Data

4.1. Data collection

To carry out the above analysis, I collect data via the YouTube Data API and via HTML webscraping. First, I use the website `channelfinder.com` to compile a list of all active German YouTube channels as of Oct 2017. Based on this list, I collect data on the YouTuber level, including a full history of video uploads by each YouTuber, from the Data API. Finally, I retrieve data on the video level, including the date of upload, video duration, views, likes, dislikes, category, and keywords. Note that views, likes and dislikes are accumulative measures; thus, I retrieve these numbers as they are on the day of data collection.

Data on the YouTubers' monetization settings is, unfortunately, highly limited; the Data API, for instance, does not provide any information regarding a video's number of ad breaks. Moreover, YouTube technically prohibits any automated program from collecting data "faster than a human could."[19] Hence, although the permitted ad breaks are detectable in a video's HTML code, a webscraper could not crawl each video in the dataset within a reasonable amount of time. Instead, I let a webscraper crawl twenty randomly drawn videos per YouTuber.[20] If it detects at least one ad break in at least one video, I classify the YouTuber as "advertising YouTuber", and as "non-advertising YouTuber" otherwise. This compromise allows me to collect monetization data on the YouTuber level for all YouTubers in my dataset, but forgoes more fine-grained information on the video level. Appendix B.1 discusses the consequences of a potential measurement error.

[19]See `www.youtube.com/static?gl=de\&template=terms\&hl=en` (Oct 2018).

[20]The webscraper pauses for eight seconds before proceeding to the next video; crawling each video this way would take several years. Crawling twenty videos per YouTuber, in contrast, is feasible within three and four months.

4.2. Definition of popular content

I use the number of video views and the videos' keywords to construct a measure for popular content. Each video is given illustrative keywords by its YouTuber – for instance, a funny cat video would be equipped with the keywords "funny" and "cat" – which help viewers to find them via YouTube's search engine.[21] For each month, for each video category, I compute how many views a certain keyword has attracted and rank them in descending order; the upper one percent of the keywords in this distribution is then classified as "popular."[22] Finally, I assign a dummy variable that is equal to one to all videos equipped with a popular keyword.[23] Note that it is important to consider each month and each video category separately. First, what is popular is likely to change over time, second, different video categories attract very different audiences whose preferences need to be considered separately. Moreover, it is crucial to define popular content based on the universe of *all* active YouTubers, i.e., before I exclude observations to construct the final dataset. Otherwise, I would compute the most popular keywords within the sample of YouTubers selected for the main analysis (see Section 4.3), which is conceptually different.

Take the category "Science & Technology" in April 2015 as an example. Videos are given 13, 555 different keywords; the three most viewed are "diy", "homemade", and "selfmade". Figure 4 shows that the distribution of views over keywords is heavily skewed: a small number of keywords accounts for a large part of the views. For instance, the upper one percent of the keywords accounts for 45.1%, while the lowest ten percent of the keywords account for just 0.02% of the views. The numbers are similar for other categories and other points in time.

4.3. Final dataset

In a last step, I construct my final dataset. First, I define an appropriate observation period. The central event – the launch of the new ad break tool – took place in Oct 2015. Including videos uploaded between Jan 2013 and Jan 2017 into the final dataset yields a sufficient number of before and after observations. At the same time, this choice excludes both videos that are too old – and therefore not well comparable to more recent ones in terms of content or duration – as well as videos that were too "recent" on the date of data collection. By leaving at least nine months between the latest upload of a video and the data collection process (that started in Oct 2017) all videos in my dataset can be considered as "old", which minimizes any potential bias that may arise through the accumulative nature of some descriptive variables such as likes, dislikes, and views. Moreover, an observation period from Jan 2013 to Jan 2017 excludes the two big demone-

[21]If a video is not given keywords, I generate keywords from its title.

[22]I ignore trivial keywords that appear in the video categories' titles. For instance, I ignore "people" and "blog" for videos in the category "People & Blogs" and "science" and "technology" for videos in the category "Science & Technology."

[23]In that, I follow the procedure by Sun and Zhu (2013), only that instead of blogs' hashtags I use videos' keywords.

tization waves from 2017 and 2018 (see Section 2) that could have affected the YouTuber's content choice. Robustness checks on my main results using other observation periods and a summary of minor events that occurred between Jan 2013 and Jan 2017 are presented in Appendix A.1 and Appendix B.2.

Second, I determine which YouTubers to include. Following the outline from Section 3, I restrict the analysis to YouTubers "left" to the ten minutes threshold before Oct 2015 (YouTubers A and B in the example), where I use a YouTuber's median video duration before Oct 2015 to define her "position" on the x-axis in Figure 3. Thus, I include only YouTubers whose median video duration before Oct 2015 is smaller than ten minutes into the final dataset. In addition, I include only YouTubers who uploaded at least one video before and after Oct 2015. Finally, due to the "adpocalypse" (see Section 2), I exclude all videos from the category "News & Politics", since many of these videos were forcefully demonetized by YouTube. The final panel dataset includes $15,877$ YouTubers with $1,349,267$ videos over a time period of 49 months. Table 1 summarizes all variables used in the main analysis. Appendix A.2 shows several robustness checks based on different selections of YouTubers.

4.4. Illustrative evidence

Based on the final dataset, this section provides illustrative evidence of the two major arguments in Section 3. That is, I confirm that the launch of the new ad break tool made the ten minutes trick more apparent and that YouTubers who were closer to the ten minutes threshold before Oct 2015 are more likely to exploit it. In addition, I provide video level evidence for an increase in the actual (not the feasible) number of ad breaks per video.

First, Figure 5 demonstrates that the advertising YouTubers' share of videos between ten and fourteen minutes increases after Oct 2015. The non-advertising YouTubers, on the other hand, are unaffected. The diverging trends confirm that the launch of the new ad break tool in Oct 2015 made the ten minutes trick more apparent to the advertising YouTubers.

In what follows, I consider only advertising YouTubers. A further comparison of advertising and non-advertising YouTubers is problematic, since they act based on entirely different motives: non-advertising YouTubers had neither chance nor interest to exploit the ten minutes trick at any point in time. Thus, I exclude the non-advertising YouTubers from my main analysis, but come back to them for falsification checks in Section 6.2.3.

As a second step, I show that (advertising) YouTubers who were closer to the ten minutes threshold before Oct 2015 are more likely to exploit it. Since "closeness" – in terms of a YouTuber's median video duration (see Section 4.3) – is a continuous measure, I cannot compare the trends of distinct groups, though. Instead, Figure 6 illustrates that the increase in YouTubers' share of videos between ten and fourteen minutes after Oct 2015 is stronger for YouTubers whose "closeness" is around the 75^{th} percentile of its distribution than for YouTubers around the 25^{th} percentile.

In addition to that, I examine the distribution of video durations before and after Oct 2015 for the same two groups of YouTubers. Figures 7 to 10 illustrate three important facts. First, if the YouTubers increase their videos' duration after Oct 2015 to benefit from YouTube's discontinuous mapping from video duration to the feasible number of ad breaks, one should see bunching just behind the ten minutes threshold *after* Oct 2015. Figures 8 and 10 show that this is the case. In addition, Figures 8 and 10 illustrate that it is appropriate to focus on the share of videos between ten and fourteen minutes: if the YouTubers exploit the ten minutes trick after Oct 2015, they start to upload videos that *just* enable them to increase the feasible number of ad breaks per video. Second, if exploiting the ten minutes threshold it is less costly for YouTubers whose median video duration was closer to the ten minutes threshold before Oct 2015, the bunching should be more pronounced for YouTubers with a higher median video duration before Oct 2015; Figures 8 and 9 confirm that this is the case. Third, the distributions of video durations before Oct 2015 in Figures 7 and 9 document that the dataset is likely to be limited to YouTubers who were ignorant of the ten minutes trick before the launch of the new ad break tool, since – in contrast to Figures 8 and 10 – the distributions of video durations are smooth around the ten minutes threshold.

I augment this illustrative evidence with the results from a formal McCrary test (McCrary, 2008), which is based on an estimator for the discontinuity at a given threshold in the density function of the running variable (H_0: no discontinuity). Here, I apply a McCrary test to obtain a measure for the discontinuity in the distributions of video durations in Figures 7 to 10. The results are displayed in Figures 11 to 14. Before Oct 2015, in Figures 11 and 13, the estimates for the discontinuities are small for both groups of YouTubers. In contrast to that, the estimate discontinuity after Oct 2015 is still small in Figure 14, but much more pronounced in Figure 12, where I consider the YouTubers whose median video duration was closer to the ten minutes threshold. Estimates and standard errors of the discontinuities can be found in Table 2.

Finally, I show that the *actual* number of permitted ad breaks in videos that are ten minutes or longer has increased. To this end, I draw a random subsample of 500 advertising YouTubers and collect *video level data* on their monetization settings (52,462 videos).[24] I consider only videos that are ten minutes or longer. I find that the average number of ad breaks in these videos has grown from 0.86 before Oct 2015 to 1.04 after Oct 2015, which corresponds to an increase of 20%. Moreover, the share of videos that has more than one ad break has increased from 17.7% to 20.7%. Finally, while 23 is the largest number of ad breaks in a single video before Oct 2015, this number has risen to 52 after Oct 2015. Thus, in the random subsample, the actual number of ad breaks has increased on the intensive as well as on the extensive margin.

[24]Collecting this fine grained data is only feasible for a small subsample of YouTubers; see Section 4.1 for details.

5. Empirical strategy

5.1. Baseline regression

This section formalizes the empirical strategy outlined in Section 3. As a first step in the empirical analysis, I define the treatment and the control group. Following the outline from Section 3, I classify a YouTuber as treated if she could increase the feasible number of ad breaks in her videos after Oct 2015. To this end, I compute each YouTuber's share of videos between ten and fourteen minutes before and after Oct 2015; if this share has increased by at least five percentage points, YouTuber i is assigned to the treatment group $(2,513$ YouTubers), and to the control group otherwise $(8,086$ YouTubers). See Appendix A.3 for robustness checks that use other classifications of the treatment and the control group.

The baseline difference-in-differences regression is given by

$$Popular_{vit} = \beta D_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \epsilon_{vit}, \tag{2}$$

where D_i indicates the treatment group, $post_t$ indicates all months after Oct 2015, X_{vit} controls for video categories, ϕ_i and ϕ_t are YouTuber and monthly fixed effects, respectively, and t_i is a YouTuber specific linear time trend. The dependent variable $Popular_{vit}$ is a dummy variable equal to one if video v of YouTuber i in month t is given a popular keyword, and zero otherwise (see Section 4.2 for details). Thus, I estimate a Linear Probability Model and the parameter β measures the average percentage point change in the probability to upload popular content for YouTubers in the treatment relative to the control group.

5.2. IV regression

5.2.1. Model

An OLS estimation of equation (2) is unlikely to yield a causal estimate of the effect of advertising on the probability to upload popular content for three interrelated reasons. First, YouTubers can self-select into the treatment group. This applies, for instance, to particularly money-loving YouTubers. If these YouTubers are at the same time more likely to upload popular content, the OLS estimate for β would be upward biased. Second, omitted YouTuber specific time-varying factors that are neither captured in the YouTuber specific linear time trend nor in YouTuber or monthly fixed effects may drive $Popular_{vit}$ and D_i at the same time. To stick with the example, some YouTubers may develop a taste for money over time. If these YouTubers are more likely to upload popular content, the OLS estimate of β would, again, be upward biased. Finally, reverse causality may generate a spurious relationship between $Popular_{vit}$ and D_i. If, for instance, YouTubers who produce more popular content are more likely and more willing to increase their number of ad breaks per video, the OLS estimate for β would be upward biased, too.

To account for the endogeneity in a YouTuber's treatment status, I use YouTubers' *median video duration before Oct 2015* as an instrument for D_i. The first stage equation is given by

$$D_i * post_t = \pi close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + u_{vit} \tag{3}$$

where $close_i$ denotes the median video duration of YouTuber i before Oct 2015. The interpretation is as follows. If $close_i$ is a valid instrument (a discussion follows in Section 5.2.3), it initiates a causal chain. As good as random variation in $close_i$ generates as good as random variation in D_i, which is isolated by the first stage. Using this exogenous variation, I can consistently estimate β in equation (2) using Two Stage Least Squares (2SLS).

The reduced form of equations (2) and (3) is given by

$$Popular_{vit} = \gamma close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + v_{vit}. \tag{4}$$

The parameters β in equation (2) and γ in equation (4) answer different questions. The parameter γ is the average effect of an additional unit of $close_i$ – i.e., an additional minute – on the difference in the probability to upload popular content before and after Oct 2015. In other words, γ measures how a better chance to increase the feasible number of ad breaks per video affects the probability to upload popular content, whereby it is comparable to an intention-to-treatment effect. In contrast to that, β is the average effect of the actual treatment status D_i on the difference in the probability to upload popular content before and after Oct 2015.

5.2.2. Instrument heterogeneity

The instrument $close_i$ is likely to affect different YouTubers in different ways. In particular, some YouTubers' treatment status may be entirely unchanged. On the one hand, some YouTubers have no interest in increasing their feasible number of ad breaks per video; these YouTubers remain untreated, no matter how close to the ten minutes threshold they are. On the other hand, some YouTubers are desperate to increase the feasible number of ad breaks per video; these YouTubers pursue the treatment, no matter how far away from the ten minutes threshold they are. Thus, the 2SLS estimate for β measures a Local Average Treatment Effect (LATE, see Angrist and Imbens, 1995), i.e., a weighted average of the individual treatment effects, where the weights capture the individual magnitudes of π_i, i.e., the extent to which $close_i$ affects $Pr(D_i = 1)$.

5.2.3. Instrument validity

The validity of $close_i$ as instrument for D_i hinges on four requirements: Instrument relevance, the exclusion restriction, instrument independence, and monotonicity. These requirements are now discussed.

Instrument relevance First, the parameter π in the first stage equation (3) must be non-zero, which means that $close_i$ must be correlated to D_i. It is plausible that the closer a YouTuber's position to the ten minutes threshold before Oct 2015, the easier it is to produce videos that are ten minutes or longer after Oct 2015, for instance, because she does not have to deviate far from her former concepts or because she does not have to spend much additional effort. Illustrative evidence is provided by Figures 6, 8, and 10 in Section 4.4. Moreover, a bivariate regression of D_i on $close_i$ yields a t-statistic of around 15. Finally, the first stage diagnostics discussed in Section 6.1 confirm the instrument's relevance.

Exclusion restriction Second, $close_i$ must operate through the single, known channel $D_i * post_t$. In other words, the instrument must not be correlated to the dependent variable $Popular_{vit}$. This is a plausible assumption, too. A YouTuber's median video duration before Oct 2015 – when the YouTubers were ignorant of the ten minutes trick's existence – is most likely a result of her personal style, taste, or preferred level of effort and orthogonal to whether the video covers popular topics or not.

The panel structure of my dataset allows me to conduct an event study that confirms the plausibility of the exclusion restriction. Based on the reduced form equation (4), I interact $close_i$ with each monthly dummy, using Oct 2015 ($t = 34$) as the baseline. This specification allows me to treat the coefficients of the interaction terms as the effect of $close_i$ on $Popular_{vit}$ relative to a base month just before the YouTubers could start to adapt their content. The event study regression equation is given by

$$Popular_{vit} = \sum_{t=1}^{33} \gamma_t close_i * pre_t + \sum_{t=35}^{49} \gamma_t close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + v_{vit}. \quad (5)$$

The interpretation of this approach is analogous to checking the validity of a parallel trends assumption. While the indirect impact of $close_i$ on $Popular_{vit}$ may accumulate over time, it should not begin before a YouTuber became aware of the new ad break tool, i.e., before the treatment status D_i was switched on. Thus, if the only way $close_i$ affects the dependent variable $Popular_{vit}$ is via $D_i * post_t$, then all estimates $\gamma_t, t \leq 33$, should be close to zero and be statistically insignificant. In contrast, the estimates $\gamma_t, t \geq 35$, should be unequal to zero and statistically significant.

Instrument independence In addition to the exclusion restriction, the instrument $close_i$ must be independent of potential outcomes and potential treatments. In other words, $close_i$ must be as good as randomly assigned such that the first stage captures the causal effect of $close_i$ on D_i. Note that reverse causality is of no concern here, because $close_i$ is by definition determined before, and D_i after Oct 2015. Yet, YouTuber specific time-varying factors that drive both $close_i$ and D_i as well as the potential manipulation of $close_i$ on behalf of the YouTubers – in the sense that they choose high values of $close_i$ to increase their treatment probability – may be an issue.

Four facts, however, argue against the manipulation of $close_i$. First, the ten minutes trick was unknown until Oct 2015. Second, YouTube did not announce the new ad break tool before its launch, so the knowledge of the ten minutes trick caught the YouTubers unprepared.[25] Third, YouTubers do not benefit from higher values of $close_i$ before Oct 2015, since the number of ad breaks per video is limited to one, irrespective of how close their are to the threshold. Finally, if a YouTuber chose a high value of $close_i$ to increase her treatment probability, she must know about the ten minutes trick; if she knew about the ten minutes trick, she would either exploit or ignore it, but she would not just move closer to the threshold.

It remains to rule out that unobserved YouTuber specific time-varying factors drive both $close_i$ and D_i. Three arguments speak against such concerns. First, t_i in equation (3) controls for YouTuber specific linear time trends; in Appendix B.3, I also include higher order polynomials of t_i into equation (3). Second, while commercial interests are a plausible driver of D_i, they are unlikely to affect $close_i$, as argued above. Third, YouTubers with a strong commercial interest might self-select into particular video categories that, in turn, require a certain video duration. The vector X_{vit} in equation (3), however, captures category specific characteristics and therefore prohibits that $close_i$ is indirectly driven by a YouTuber's commercial interest.

Monotonicity Finally, while $close_i$ may have no effect on some YouTubers (see Section 5.2.2), those who are affected must be affected in the same direction, i.e., $\pi_i \geq 0 \; \forall \; i$. Again, this is a plausible assumption: it is hard to believe that a high value of $close_i$ prohibits treatment from YouTubers who would have been treated if $close_i$ was low. Figure 15 provides illustrative evidence. It plots all values of $close_i$ against the corresponding probability of treatment, $Pr(D_i = 1)$. With the exception of some outliers at the upper left and the lower right corner, the relationship between $close_i$ and $Pr(D_i = 1)$ is monotone.

Note that I might violate the monotonicity assumption if I used a continuous measure of treatment intensity – i.e., the extent to which a YouTuber increases her share of videos that are ten minutes or longer – instead of the binary treatment status D_i. As argued, YouTubers with high values of $close_i$ have a higher *probability* to increase their share of videos that are ten minutes or longer. At the same time, however, they have *less scope* to do so, because their initial share of videos that are ten minutes or longer is already high. Hence, while the impact of $close_i$ on the *extensive margin* of treatment is monotone and increasing – as shown above – it might follow an inverted U-shape on the *intensive margin*.

5.2.4. Additional requirements

In addition to the validity of the instrument, two further requirements must be fulfilled. First, to be consistent with the idea of the identification strategy, the effect of an increase

[25] I searched through the YouTube creators blog (`https://youtube-creators.googleblog.com/`) and found no entries announcing the new ad break tool from before Oct 2015.

in the feasible number of ad breaks per video on the probability to upload popular content must be driven by the videos that are ten minutes or longer. Second, video duration as such must not have a direct impact on the probability to upload popular content. These additional requirements are now discussed.

Evidence from the video level The parameter β in equation (2) aggregates the effect of an increase in the feasible number of ad breaks on the probability to upload popular content on the YouTuber level. The aggregation is coherent with my empirical strategy: the instrument provides as good as random variation on the YouTuber level, too. Yet, to check if the aggregate effect is driven by videos that are ten minutes or longer, I augment the first stage regression equation (3) to

$$I(\geq 10)_{vit} = \alpha close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \upsilon_{vit} | Popular_{vit}, \tag{6}$$

where $I(\geq 10)_{vit}$ indicates if video v of YouTuber i in month t is ten minutes or longer. Then, I estimate equation (6) by OLS for popular and for non-popular content separately.

The interpretation is as follows. The parameter α measures the effect of an additional unit of $close_i$ on the probability that a video is ten minutes or longer, conditional on whether the video is popular or not. Suppose β in equation (2) is negative, i.e., an (aggregate) increase in the feasible number of ad breaks per videos reduces the probability to upload popular content. If the aggregate effect is driven by the videos that are ten minutes or longer, the OLS estimate for α should be positive and statistically significant when I condition on non-popular content, because the probability that a video is ten minutes or longer increases within this subsample. In contrast to that, the OLS estimate for α should be close to zero and not statistically significant when I condition on popular content. Note that reverse causality concerns prohibit an interaction of the term $close_i * post_t$ in equation (6) with a dummy that indicates popular content and a corresponding regression based on the entire sample. If, for instance, an increase in the feasible number of ad breaks led to a reduction in the probability to upload popular content, the estimate for the triple interaction would be downward biased.

Video duration and popular content Finally, when a YouTuber is treated, not only her treatment status D_i changes, but – by construction – her videos' duration increases, too. Hence, I must also ensure that video duration as such does not affect the dependent variable $Popular_{vit}$.

The difficulty resembles a regression discontinuity design: when comparing observations left and right to a cutoff, not only the treatment status, but also the value of the assignment variable determining the treatment status changes. Standard regression discontinuity designs would include the assignment variable as a control. Simply controlling for video duration may, however, be problematic in my application, since the videos' duration after Oct 2015 may be manipulated to exploit the ten minutes trick. In contrast to that, the videos' duration before Oct 2015 is – similar to the instrument $close_i$ – likely to be as good as randomly assigned. Hence, I run the following regression

$$Popular_{vit} = \delta duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \epsilon_{vit} \mid t \leq 33 \tag{7}$$

including only the time period before Oct 2015, where I expect δ to be close to zero and statistically insignificant.

6. Results

6.1. Main results

Table 3 presents the main results. Columns 1 to 3 show the results from the potentially biased OLS estimation of equation (2). The estimates are close to zero and not statistically significant despite the large sample size. In contrast to that, the estimates obtained by a 2SLS estimation of equations (2) and (3), displayed in columns 4 to 6, are negative and statistically significant at the 1%-level. According to these estimates, an increase in the feasible number of ad breaks per video decreases the probability to duplicate popular content by about twenty percentage points. The effect size is considerable: it corresponds to 40% of a standard deviation in the dependent variable $Popular_{vit}$ and to around 50% of its baseline value 0.448. The large difference between the OLS and the 2SLS estimates confirms the endogeneity concerns expressed earlier: YouTubers' self-selection into treatment, omitted YouTuber specific time-varying factors as well as reverse causality may lead to an upward bias in the estimate for β when not taken into account.

The first stage diagnostics in columns 4 to 6 confirm the validity of my empirical strategy. Having been closer to the ten minutes threshold before Oct 2015 leads to a higher treatment probability: an additional unit of $close_i$ (i.e., an additional minute) increases the treatment probability by about 2.9 percentage points. The estimate is highly statistically significant. Moreover, an F-statistic between 144 and 151 demonstrates the strength of the instrument (Stock and Yogo, 2002; Kleibergen and Paap, 2006).

Finally, columns 7 to 9 show the reduced form estimates of equation (4). As argued in Section 5, these estimates measure the effect of an additional unit of $close_i$ on the probability to duplicate popular content. Consistent with the results from the 2SLS regression, the estimates are negative: a one unit increase in $close_i$ leads to a 0.6 percentage point reduction in the probability to duplicate popular content. Though small, the estimates are statistically significant at the 1%-level.

In sum, the results presented in Table 3 lead to the conclusion that the exogenous increase in the feasible number of ad breaks per video causes a considerable reduction in the probability to duplicate popular content. In other words, I find evidence that advertising has a causal positive effect on content differentiation. The results match the theoretical considerations discussed in Section 1: when the YouTubers increase the number of ad breaks in their videos, they raise the ad "price" that their viewers have to pay. A higher ad "price", in turn, goes along with higher content differentiation. A detailed discussion of the economic mechanism follows in Section 7. Appendix A.5 shows that the main results are robust to alternative measures of popular content.

6.2. Validity checks

6.2.1. Exclusion restriction

This section confirms the plausibility of the exclusion restriction as discussed in Section 5.2.3. Figure 16 presents the results of the event study. The dots connected by the solid line display the estimates γ_t from a regression of equation (5), the dashed lines depict a 95% confidence interval. The estimates for γ_t, $t \in [1, 33]$, fluctuate around zero without a visible trend. The lion's share of the estimates is not statistically significant at the 5%-level. In contrast to that, the estimates for γ_t, $t \in [35, 49]$, are negative and downward trending. Moreover, most estimates are statistically significant at the 5%-level. Hence, $close_i$ had no clear and statistically significant effect on the dependent variable $Popular_{vit}$ before Oct 2015, but a clear negative and increasingly strong effect after Oct 2015. See Appendix A.6 for a series of placebo regressions that supports the plausibility of the exclusion restriction, too.

6.2.2. Additional requirements

Next, I show that the additional requirements from Section 5.2.4 are fulfilled. First, Table 4 displays the results from an OLS regression of equation (6) on the subsample of popular videos (columns 1 to 3) and on the subsample of non-popular videos (columns 4 to 6). While the OLS estimate of α is small and not statistically significant in columns 1 to 3, it is around six times larger and statistically significant at the 5%-level in columns 4 to 6. These results are consistent with the ideas from Section 5.2.4. If the estimate for β is negative and if this aggregate effect is driven by the videos that are ten minutes or longer, the estimate for α should be close to zero when considering only popular, and positive when considering only non-popular content.

Second, I consider the regression results from equation (7). Table 5 shows that the estimate for δ in equation (7) is very small and statistically insignificant. Thus, I find no evidence in my data that video duration as such directly affects $Popular_{vit}$.

6.2.3. Non-advertising YouTubers

The non-advertising YouTubers, whom I do not consider in the main analysis, allow me to conduct two additional validity checks. The non-advertising YouTubers' content choices are not driven by commercial considerations. As a consequence, their probability to upload popular content cannot be affected by the launch of the new ad break tool in Oct 2015. Hence, non-advertising YouTubers who are classified as "treated" must have increased their share of videos between ten and fourteen minutes for reasons other than exploiting the ten minutes trick. As a consequence, the estimate for β should be close to zero and statistically insignificant when I estimate equations (2) and (3) by 2SLS on the non-advertising YouTubers only.

The regression results in Table 6 support these considerations. While the potentially biased OLS estimates in columns 1 to 3 are positive and significant at the 5%-level, both the IV estimates (columns 4 to 6), and the reduced form estimates (columns 7 to 9) are close to zero and statistically insignificant.

Figure 17 provides an additional plausibility check of the exclusion restriction. If the only way the instrument $close_i$ affects the dependent variable $Popular_{vit}$ is via the increase of the feasible number of ad breaks per video, then *all* estimates γ_t obtained when estimating equation (5) on the subsample of non-advertising YouTubers should be close to zero and be statistically insignificant. Figure 17 demonstrates that this is the case.

6.3. Effect heterogeneity

One particular strength of my dataset is its size, which allows me to conduct a series of subgroup analyses. To this end, this sections illustrates effect heterogeneity along two dimensions. First, I show that the average effect from Section 6.1 is driven by YouTubers with many subscribers. Second, I document that some video categories are more flexible regarding their typical video duration, which leads to heterogeneity on the first stage.

6.3.1. Heterogeneity along the subscriber count

The adaption of video content entails costs. The YouTubers must deviate from the content they were producing before, which may force them to focus on topics that they are less intrinsically motivated to cover. The larger a YouTuber's audience, however, the higher is her benefit from additional ad breaks and therefore also the probability that the additional ad revenue covers the costs. To confirm that the effect of an increase in the feasible number of ad breaks on the probability to upload popular content is stronger for YouTubers with a high subscriber count, I split my sample at the $1,000$ subscriber threshold – which roughly corresponds to the median number of subscribers – and consider YouTubers with at least $1,000$, and YouTubers with fewer than $1,000$ subscribers separately.[26] Note that reverse causality prohibits including the subscriber count as an interaction term. If, for instance, YouTubers who upload much popular content have a larger audience, I would overestimate the effect of a YouTuber's subscriber count.

Tables 7 and 8 show the results from regressing equations (2) and (3) on the two subsamples. The potentially biased OLS estimates in columns 1 to 3 are close to zero and statistically insignificant in both tables. The 2SLS estimates, however, are larger than the average effect in Table 3 when considering the YouTubers with at least $1,000$ subscribers,

[26] YouTube has also recently disabled all YouTube channels with fewer than $1,000$ subscribers from monetization, arguing that this is a meaningful threshold for a channel to be considered "eligible" for ad revenues (see Section 2 for details).

but close to zero for the other subsample. The first stage estimates follow a similar pattern: they are around 15% smaller than in Table 3 when considering YouTubers with fewer than $1,000$ subscribers, but statistically significant at the 1%-level in both cases. Finally, consistent with the 2SLS results, the reduced form estimates in columns 7 to 9 are larger than the average effect and statistically significant for the YouTubers with at least $1,000$ subscribers in Table 7, but close to zero for the YouTubers with fewer than $1,000$ subscribers in Table 8. Thus, while an increase in the feasible number of ad breaks leads to an increase in content differentiation for YouTubers with a relatively large audience, YouTubers with a low subscriber count refrain from adapting their content.

6.3.2. Heterogeneity along video categories

Next, I demonstrate that some video categories are more flexible regarding their typical video duration. For instance, a music clip typically takes between three and five minutes and cannot be easily extended to ten minutes. Similarly, a comedy video becomes boring if it does not get the gag across. To illustrate heterogeneity between the fourteen video categories considered in the analysis, I estimate equations (2) and (3) on fourteen subsamples that include all videos from a particular video category.

The results in Table 9 reveal effect heterogeneity in terms of the first and also in terms of the second stage. The first stage estimate is close to zero for the categories "Music", "Comedy", and "Let's Play." Let's Play videos are often based on how YouTubers finish video game levels, many of which include a time constraint. The first stage estimate is largest for the categories "Cars & Vehicles", "Pets & Animals", and "Sports", hence, videos from these categories can either be most easily extended to ten minutes or more, or YouTubers who have the strongest desire to increase their feasible number of ad breaks self-select into these categories. The first stage estimate for the remaining categories is similar to the results from Section 6.1.

For the discussion of the second stage estimates, I focus on the categories with a first stage F-statistic above 10. Consistent with the main results from Section 6.1, all estimates are negative; their size ranges from -0.0762 ("Cars & Vehicles") to -0.922 ("Film & Animation"). The estimates are statistically significant for the categories "Film & Animation", "People & Blogs", and "Entertainment", which are also the categories with the highest number of observations. Hence, in addition to heterogeneity on the first stage, the video categories differ in the extent to which the video content is adapted. There are, again, two plausible explanations. First, it could be easier to create videos that cover non-popular content for some categories; in other words, the effect heterogeneity is driven by category specific differences (that are not captured by X_{vit}). Second, YouTubers who are more creative or more willing to try out something new might self-select into the video categories "Film & Animation", "People & Blogs", and "Entertainment" whose second stage effect is strongest.

6.4. Differentiation along the tail

Up to this point, I have considered the effect of an increase in the feasible number of ad breaks per video on the *most* popular content only. In this section, I study content differentiation along the "tail." In particular, I show that the effect I document in Section 6.1 diminishes for less popular content until it eventually switches its sign. To this end, I generate five dummy variables that indicate alternative percentiles of the distribution of most-viewed keywords (see Section 4.2 for details): the 1^{st} to 10^{th}, the 10^{th} to 25^{th}, the 25^{th} to 50^{th}, the 50^{th} to 75^{th}, and the 75^{th} to 100^{th} percentile. Then, I replace the dependent variable $Popular_{vit}$ in equation (2) with each of these dummies and estimate equations (2) and (3) by 2SLS.

The results in Table 10 illustrate the pattern. The estimate for β in column 1 is similar to its counterpart in Table 3: an increase in the feasible number of ad breaks per video leads to a 20% percentage point reduction in the probability to upload a video that is given a keyword from the 1^{st} to 10^{th} percentile of the distribution of most-viewed keywords. The effect size corresponds to nearly 50% of a standard deviation in the dependent variable. The estimate, however, decreases by half in columns 2 and 3, and by about two-thirds in column 4. Finally, in column 5, the estimate switches its sign and becomes positive. The effect size, however, is small: it corresponds to 15% of a standard deviation in the dependent variable. All estimates for β are statistically significant.

To interpret these results, note that a video is given around eleven keywords on average and that this number is constant over time. Hence, a video can be given keywords from several parts of the distribution of most-viewed keywords. Bearing this mind, the estimates in Table 10 demonstrate that the YouTubers who could increase the feasible number of ad breaks per video do not move from exclusively uploading popular only to uploading non-popular content only. Rather, they change the "mixture" of topics in a video: they abandon covering popular topics – the more popular, the stronger the effect – and cover more of the less popular topics instead. Thereby, the probability to upload very popular content decreases, while the probability to upload not so popular content remains unchanged or increases only slightly. Indeed, when I count each video's number of "affiliations" to the categories displayed in Table 10 and use this count as dependent variable in equation (2), a 2SLS estimation shows that videos from YouTubers who could increase the feasible number of ad breaks per video are given keywords from fewer different categories after Oct 2015 than before (column 6 in Table 10).

7. Mechanism

This section studies the economic mechanism that drives the results from Section 6. In particular, I show that YouTubers who increase the feasible number of ad breaks per video avoid competition in the ad "price": since popular content is also covered by many YouTubers, viewers could easily switch to a different channel if a YouTuber increased her ad "price." Switching becomes less likely, however, when the YouTuber uploads content that is less popular and thereby covered by fewer competitors. I define a measure for "competitive content", i.e., a measure for the most-covered content on YouTube, and show that it is highly correlated to popular content. Then, I demonstrate that an increase in the feasible number of ad breaks per video reduces the YouTubers' probability to upload competitive content. Since competitive content is typically also popular – i.e., content in high demanded is also supplied by many YouTubers – competition in the ad "price" ultimately leads to a reduction in the probability to upload popular content. Finally, I support this result by demonstrating that the audience of YouTubers who could increase the feasible number of ad breaks per video becomes more stable, i.e., the viewers become less likely to switch to competitors. In contrast to that, I find no evidence for a YouTuber learning effect (see Appendix B.3).

7.1. Definition of competitive content

First, I construct a measure of "competitive content", i.e., a measure for the most-covered content on YouTube. The procedure is analogous to the definition of "popular content" (see Section 4.2). For each month, for each video category, I compute how many times a certain keyword has been *used* and rank them in descending order; the upper one percent of this distribution is classified as "competitive." Then, I assign a dummy variable that is equal to one to all videos equipped with a competitive keyword. Note that a competitive keyword is not necessarily a popular keyword, too. A keyword may attract many views although it is not used by many YouTubers; similarly, a keyword may be used by many YouTubers, but does not attract many views. In my sample, the correlation between popular and competitive content is equal to 0.57.

Take the category "Science & Technology" in April 2015 as an example again. The three most used keywords are "deutsch", "test", and "review" (note that they are different from the three most viewed keywords, see Section 4.2). Figure 18 shows that the distribution of usages over keywords is heavily skewed. For instance, the upper one percent of keywords accounts for 17.4%, while the lowest ten percent account for 4.4% of all keyword usages.[27] The numbers are similar for other categories and other points in time.

7.2. IV regression

Analogous to Section 5, the baseline regression equation is given by

$$Competitive_{vit} = \beta' D_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \epsilon_{vit}, \tag{8}$$

where the dependent variable $Competitive_{vit}$ is equal to one if video v of YouTuber i in month t is given a competitive keyword as defined above. Thus, I estimate a Linear Probability Model, where the parameter β' measures the average percentage point change

[27]Many keywords are used rarely – e.g., once or twice – which is responsible for the steps in the plot.

in the probability to upload competitive content for YouTubers in the treatment relative to the control group.

As for equation (2), an OLS estimation of equation (8) is unlikely to yield the causal effect of advertising on the probability to upload competitive content for three interrelated reasons (see Section 5.2.1 for a detailed discussion of these concerns). First, YouTubers can self-select into the treatment group. Second, omitted YouTuber specific time-varying factors might drive $Competitive_{vit}$ and D_i at the same time. Third, reverse causality may be an issue. To account for the endogeneity in a YouTuber's treatment status, I use equation (3) as a first stage again and estimate equations (3) and (8) by 2SLS.

Finally, the reduced form of equations (3) and (8) is given by

$$Competitive_{vit} = \gamma' close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + v_{vit}, \tag{9}$$

where γ' measures the effect of an additional unit of $close_i$ on the probability to upload competitive content.

7.3. Results

7.3.1. YouTubers avoid competition

The estimates for β' in Table 11 are similar to the estimates for β in Table 3. Columns 1 to 3 in Table 11 show the results from a potentially biased OLS estimation of equation (8). The estimates are close to zero and not statistically significant. In contrast to that, the estimates obtained by a 2SLS estimation of equations (3) and (8), displayed in columns 4 to 6, are negative and statistically significant at the 1%-level. According to these estimates, an increase in the feasible number of ad breaks per video decreases the probability to upload competitive content by about twenty percentage points; the effect size corresponds to 42% of a standard deviation in the dependent variable $Competitive_{vit}$ and to around 30% of its baseline value 0.65. As in Section 6.1, the large difference between the OLS and the 2SLS estimates confirms the endogeneity concerns about a YouTuber's treatment status D_i. Finally, columns 7 to 9 show the reduced form estimates of equation (9). Consistent with the results from the 2SLS estimation, the estimates are negative: a one unit increase in $close_i$ leads to a 0.6 percentage point reduction in the probability to upload competitive content. Though small, the estimates are significant at the 1%-level.

The results in Table 11 show that an increase in the feasible number of ad breaks per video reduces the YouTubers' probability to upload competitive content. Given that the dependent variables $Popular_{vit}$ and $Competitive_{vit}$ are highly correlated, this is no surprise. Thus, the estimates confirm that YouTubers who increase their ad "price" avoid competition over competitive content, which is a plausible economic mechanism that drives the results from Section 6.

7.3.2. Validity checks

Although the measures are highly correlated, "competitive content" is conceptually different from "popular content." This section conducts three validity checks to show that the empirical strategy from Section 5 is also valid when I use $Competitive_{vit}$ as dependent variable in equation (8). First, I confirm the plausibility of the exclusion restriction. Second, I show that the effect of an increase in the feasible number of ad breaks per video on the probability to upload competitive content is driven by videos that are ten minutes or longer. Finally, I rule out that video duration as such has a direct effect on the probability to upload competitive content. See Appendix A.7 for further robustness checks.

Exclusion restriction To show that the instrument $close_i$ has no direct effect on the dependent variable $Competitive_{vit}$, I conduct an event study as outlined in Section 5.2.3. Based on the reduced form regression equation (9), I interact $close_i$ with each monthly dummy, using Oct 2015 ($t = 34$) as the baseline.

Figure 19 shows the results. The estimates for γ'_t, $t \in [1, 33]$, fluctuate around zero without a visible trend; the lion's share of the estimates is not statistically significant at the 5%-level. In contrast to that, the estimates for γ'_t, $t \in [35, 49]$, are negative with a downwards trend, and most of them are statistically significant. Thus, while $close_i$ has no clear and statistically significant effect on $Competitive_{vit}$ before Oct 2015, its impact is negative and increasingly strong after Oct 2015, when it could operate through the treatment status D_i.

As a further plausibility check, I conduct the event study on the subsample of non-advertising YouTubers. If $close_i$ affects $Competitive_{vit}$ only through an increase of the feasible number of ad breaks per video, then *all* estimates for γ'_t should be close to zero and statistically insignificant when considering the non-advertising YouTubers only. Figure 20 shows that this is the case. Although there is a downward trend in the estimates after Oct 2015, about a third of them has a positive sign, all of them are small, and no estimate is statistically significant at the 5%-level.

Evidence from the video level Similar to β, the parameter β' in equation (8) aggregates the effect of an increase in the feasible number of ad breaks on the probability to upload competitive content on the YouTuber level. Analogous to the approach from Section 5.2.4, I estimate

$$I(\geq 10)_{vit} = \alpha' close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \upsilon_{vit} | Competitive_{vit}, \tag{10}$$

by OLS for competitive and non-competitive content separately. If the aggregate effect from Table 11 is driven by the videos that are ten minutes or longer, the OLS estimate for α' should be positive and statistically significant when I condition on non-competitive content, but close to zero for competitive content (see Section 5.2.4 for a discussion). The results in Table 12 show that this is the case: the estimate for α' is close to zero and not statistically significant for competitive (columns 1 to 3), but several times as large and significant at the 5%-level for non-competitive content (columns 4 to 6).

Video duration and competitive content To check if video duration as such has no effect on the probability to upload competitive content, I estimate

$$Competitive_{vit} = \delta' duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \epsilon_{vit} \mid t \leq 33 \tag{11}$$

by OLS, including only observations from before Oct 2015. Table 13 shows that the estimate for δ' is close to zero and not statistically significant (see Section 5.2.4 for a discussion). Thus, I find no evidence in my data that video duration as such directly affects $Competitive_{vit}$.

7.4. Commentator fluctuation

Finally, I provide evidence of a decrease in the viewer fluctuation of YouTubers who could increase the feasible number of ad breaks per video. If a YouTuber uploads less popular content, she decreases the probability that viewers switch to competitors, because the video supply is lower. Thus, the YouTuber's viewer fluctuation should go down, which means that a given number of views should be generated by a smaller number of different viewers than before.

Since data on a YouTuber's viewers is not available, I use her videos' commentators as a proxy variable and define YouTuber i's commentator fluctuation as

$$fluctuation_i = \frac{commentators_i}{comments_i}, \tag{12}$$

where the numerator refers to the number of unique commentators of YouTuber i and the denominator refers to the total number of comments she received. If each comment on i's videos is left by a different commentator, $fluctuation_i$ is equal to 1. If several comments are written by the same commentator, $fluctuation_i$ is smaller than 1. Finally, if YouTuber i never receives any comment, $fluctuation_i$ is not defined.

Next, I compute each YouTuber's *change* in $fluctuation_i$ before and after Oct 2015,

$$\Delta fluctuation_i = fluctuation_{i,post} - fluctuation_{i,pre}, \tag{13}$$

where $fluctuation_{i,post}$ is based on the fifteen months after, and $fluctuation_{i,pre}$ is based on the fifteen months before and including Oct 2015.[28] A decrease in YouTuber i's commentator fluctuation after Oct 2015 implies that $\Delta fluctuation_i < 0$; an increase implies that $\Delta fluctuation_i > 0$.

To check if an increase in the feasible number of ad breaks leads to a decrease in the

[28]Since I have 34 observation periods before and including Oct 2015, but only fifteen observation periods afterwards, I restrict the computation of $fluctuation_{i,pre}$ to the fifteen most recent ones to increase the comparability to $fluctuation_{i,post}$.

YouTubers' commentator fluctuation, I use $\Delta fluctuation_i$ as dependent variable in the regression equation

$$\Delta fluctuation_i = \rho_0 + \rho_1 D_i + \epsilon_i, \tag{14}$$

where ρ_1 measures how the YouTubers' treatment status D_i affects their average change in commentator fluctuation. To account for endogeneity in D_i, I use

$$D_i = \psi_0 + \psi_1 close_i + e_i, \tag{15}$$

as a first stage and estimate equations (14) and (15) by 2SLS (see Section 5.2.1 for a detailed discussion of the endogeneity concerns). Since $fluctuation_i$ is sensitive to additional commentators when the total number of comments is small – for instance, if a YouTuber has only received three comments, it makes a big difference if they are written by two or three different commentators – I restrict the analysis to YouTubers who received at least 25 comments before and after Oct 2015 (see Appendix A.7.3 for alternative thresholds).

Table 14 shows the results. Column 1 presents the potentially biased OLS estimate for ρ_1. The estimate is negative: an increase in the feasible number of ad breaks per video leads to a decrease in $\Delta fluctuation_i$. The 2SLS estimate in column 2 is negative, too, but more than three times larger in absolute value than the OLS estimate. The effect size XXX. The reduced form estimate in column 3 is consistent with the 2SLS estimate in column 2. All estimates are statistically significant. In contrast to that, the 2SLS and reduced form estimates are not statistically significant when I conduct the same analysis on the subsample of non-advertising YouTubers (Table 15). Thus, I find that an increase in the feasible number of ad breaks leads to a decrease in the YouTubers' commentator fluctuation, which supports the plausibility of the results from Section 7.3 along with the argument that YouTubers upload less popular content to avoid competition in the ad "price."

8. Differentiation in the aggregate

Up to this point, I have studied how advertising affects *individual* YouTubers' content choice. In contrast to that, this section shows how content differentiation develops *in the aggregate*. In particular, I study if the tail of keywords becomes "longer" (i.e., if the total number of keywords increases), and if the tail becomes "fatter" (i.e., if the concentration of videos on keywords decreases). I do not make causal claims here; rather, I pursue a descriptive before-after comparison to put the results from Sections 6 and 7 into a broader context.

I have two options to analyze content differentiation in the aggregate: I could continue to focus on the subsample of YouTubers whom I selected for the main analysis in Section 4.3 or I could examine the entire population of German YouTubers. Proceeding with the subsample has the advantage of computing aggregate measures that are solely based on YouTubers who have the option to increase their feasible number of ad breaks per video, but would not reveal how the *entire* video supply on YouTube develops after Oct 2015. Even YouTubers who are not directly affected by the launch of the new ad break tool may adapt their video content as a reaction to their competitors' change in content; thus, studying the population of YouTubers might be more informative about aggregate developments. On the other hand, the content choices of YouTubers whom I did not select for the main analysis could be driven by motives that are orthogonal to the launch of the new ad break tool and its consequences; such effects might superimpose the treatment's aggregate effect on content differentiation and complicate the interpretation of the net effect. Since no approach clearly excels the other, I pursue both options and interpret the results adequately.

8.1. The tail becomes longer

To show that the tail of keywords becomes longer both within the subsample and the entire population of YouTubers, I compute the absolute number of unique keywords before and after Oct 2015. As I observe 34 months before (and including) Oct 2015, but only 15 months afterwards, I limit the analysis to the 15 most recent months before (and including) Oct 2015.

In the subsample, there exist $607,358$ unique keywords before, and $875,503$ unique keywords after Oct 2015, which corresponds to an absolute increase of $268,145$ unique keywords and to a relative increase of 44.15%. Considering the population of YouTubers, I find that there exist $1,090,355$ unique keywords before, and $2,096,373$ unique keywords after Oct 2015, which corresponds to an absolute increase of $1,006,018$ keywords and to a relative increase of 92.27%. The results match the findings from Sections 6 and 7: it is plausible that the total number of unique keywords increases when the YouTubers reduce the probability to upload popular or competitive content. The difference in the results could stem from entry: by construction, the population includes all YouTubers who entered the platform after Oct 2015, which may further increase the number of unique keywords that exist after Oct 2015.

8.2. The tail does not become fatter

To study if the tail becomes "fatter", I compute a Gini coefficient for the concentration of videos on keywords before and after Oct 2015.[29] Again, I restrict the analysis to the 15 most recent months before (and including) Oct 2015. Note that the Gini coefficient for the subsample measures the concentration of videos on keywords that occur *within the subsample*, while the Gini for the population measures the concentration of *all* videos on *all* keywords.

[29] I.e., the keywords replace the households, and the number of videos that use a certain keyword replaces the income in a conventional Gini computation. Note, also, that I do not use absolute measures of concentration such as the Herfindahl index, because the number of keywords before and after Oct 2015 is different.

The Gini coefficient for the subsample is high and remains nearly unchanged: is equal to 0.800 before, and equal to 0.806 after Oct 2015, which corresponds to an increase of 0.75%. Thus, the YouTubers whom I initially selected for the main analysis do not differentiate from each other after Oct 2015. The result does not contradict the findings from Section 7, though. My measures for popular and competitive content are based on *all* active German YouTubers. It is therefore possible that the YouTubers in the subsample decrease their probability to upload competitive content, where competitive content takes *the population of* YouTubers into account, but that the concentration of videos on keywords *within* the subsample remains nearly unchanged. In addition to that, the tail of keywords becomes longer after Oct 2015 (see Section 8.1). If many of those additional keywords are used by a small number of videos, the Gini coefficient as a relative measure of concentration remains unchanged even if the concentration of videos on the remaining keywords decreases.

The Gini coefficient for the population of German YouTubers increases from 0.848 before to 0.862 after Oct 2015, which corresponds to an increase of 1.65%. Here, too, the increase in the relative concentration measure could be due to the large amount of additional keywords. It is also possible that further developments – orthogonal to the launch of the new ad break tool – superimpose the effect of an increase in the feasible number of ad breaks on content differentiation in the aggregate. For instance, the growing popularity of the platform may have led to a large number of entrants who copy from the most popular YouTubers and thereby increase the concentration of videos on keywords.

9. Quality

As an extension of the main analysis, this section studies the effect of an increase in the feasible number of ad breaks on video quality. Two predictions compete. On the one hand, a higher number of ad breaks per video implies that each viewer is c.p. more valuable than before; hence, the incentive to provide high quality goes up. In addition, the YouTubers may want to counterbalance their viewers' increased ad nuisance costs. On the other hand, YouTubers could not only avoid competition in the ad "price", but also competition in terms of video quality when they reduce their probability to upload competitive content (see, e.g., Bourreau, 2003; Armstrong and Weeds, 2007; Weeds, 2013); as a result, the incentive to provide high quality diminishes. Moreover, YouTubers deviate from the content they were providing before and which they might have been more intrinsically motivated to cover. A lack of passion could have a negative effect on their videos' quality (see Sun and Zhu, 2013, for a similar argument). The results from two different measurement approaches are ambiguous: while an increase in the feasible number of ad breaks per video leads to a decrease in the fraction of likes, it leads to an increase in the number of views.

9.1. Likes and dislikes

First, I use a video's number of likes and dislikes to measure its quality. To this end, I normalize the number of likes of video v by YouTuber i in month t by its sum of likes and dislikes: $\frac{Likes}{Likes+Dislikes}_{vit}$. Though straightforward to interpret, this measure reflects the viewers' general satisfaction with a video, which is determined by its quality *and* the viewers' ad aversion. Thus, even if an increase in the feasible number of ad breaks led to an increase in video quality, a video's fraction of likes could decrease if the viewers' additional ad nuisance costs prevail.

I replace the dependent variable $Popular_{vit}$ in equation (2) with $\frac{Likes}{Likes+Dislikes}_{vit}$ and estimate equations (2) and (3) by 2SLS. Table 16 shows the results. Again, the potentially biased OLS estimates of equation (2) in columns 1 to 3 are close to zero and not statistically significant. In contrast to that, the 2SLS estimates in columns 4 to 6 are negative and statistically significant at the 1% level: an increase in the feasible number of ad breaks leads to a 4 percentage point reduction in the fraction of likes. The effect size corresponds to around 25% of a standard deviation in the dependent variable $\frac{Likes}{Likes+Dislikes}_{vit}$ and to 4.4% of its baseline value 0.91. The reduced form estimates in columns 7 to 9 are in line with these results. Note that I lose $77,066$ videos that have not received any likes or dislikes.

The results in Table 16 illustrate that viewer satisfaction has gone down. It is, however, unclear if the effect is driven by a decrease in video quality or by the viewers' irritation from additional ad breaks. See Appendix A.8 for validity checks.

9.2. Views

Second, viewers "vote with their feet." Hence, I use a video's number of views as a further measure of quality. YouTube counts a view if the video is watched for at least thirty seconds; if the video is shorter than that, it must be watched entirely.[30] If an increase in the feasible number of ad breaks led to an increase in video quality, more viewers may watch the video for more than thirty seconds. In addition, more viewers may watch the video repeatedly.

Analogous to Section 9.1, I replace the dependent variable $Popular_{vit}$ in equation (2) with the logarithm of the number of views of video v by YouTuber i in month t: $log(Views)_{vit}$. Then, I estimate equations (2) and (3) by 2SLS. Table 17 shows the results. The potentially biased OLS estimates in columns 1 to 3 are positive and statistically significant at the 1%-level. According to these estimates, an increase in the feasible number of ad breaks leads to a 20% increase in views. The 2SLS estimates in columns 4 and 5 are more than twice as large and statistically significant at the 1%-level, too. The 2SLS estimate is, however, sensitive to including a YouTuber specific linear time trend: in column 6, it diminishes by about a third relative to columns 4 and 5. Moreover, the estimate is only weakly statistically significant at the 10%-level. The reduced form estimates match

[30] See www.tubics.com/blog/what-counts-as-a-view-on-youtube/ (May 2019).

the pattern. They are positive and statistically significant at the 1%-level in columns 7 and 8, but only at the 5%-level when I add a YouTuber specific linear time trend in column 9.

There are two potential explanations for the differences to Section 9.1. First, video quality may enhance, whereby more (repeated) viewers are attracted. At the same time, however, viewers express their dissatisfaction with the additional breaks by a disliking the video. Second, there could be algorithmic confounding of the data (Salganik, 2017, Ch. 3). YouTube, too, earns a fraction of the YouTubers' ad revenue. Thus, the platform has an incentive to treat videos with many ad breaks favorably, for instance, through its ranking algorithm. In this case, the number of views was not informative about a video's quality, but only about an algorithmic advantage. See Appendix A.8 for validity checks.

10. Conclusion

This paper demonstrates that an increase in the feasible number of ad breaks per video leads to an increase in content differentiation between several thousand YouTubers. In particular, I find that an increase in the feasible number of ad breaks per video reduces the YouTubers' probability to duplicate popular content by about twenty percentage points, because YouTubers avoid competition in the ad "price." The results provide empirical evidence for predictions from economic theory: models that acknowledge the conceptual equivalence between direct prices and consumers' nuisance costs from advertising find that media outlets prefer to differentiate from each other to avoid ruinous competition in the ad "price."

The paper advances debates on the effect of advertising on content differentiation. In particular, showing that advertising does *not* lead to the duplication of popular content entails two implications for present policies. First, advertising quantities are often restricted in an attempt to protect consumers.[31] The Audiovisual Media Services Directive, for instance, requires that the proportion of television advertising and teleshopping spots within a given clock hour shall not exceed 20% (Article 23 §1). My paper demonstrates that consumers may *benefit* from advertising, because it increases content differentiation; policy makers need to take this additional effect into account when they determine advertising quantity restrictions. Similarly, public interventions in television markets – i.e., public service broadcasters – grow from the claim that advertising funded broadcasting fails to serve all viewers' preferences over content (Armstrong and Weeds, 2007). My results controvert this argument: advertising leads to *more* content differentiation. Thus, while valuable contributions to culture, education, and the public discourse certainly justify public service broadcasting, concerns about content duplication by advertising funded broadcasters do not.

My paper is limited in at least four respects. First, although I present competition in the ad "price" as a plausible mechanism for my main results and rule out a YouTuber

[31] See `www.ofcom.org.uk/__data/assets/pdf_file/0021/19083/advertising_minutage.pdf` (Dec 2018).

learning effect, I cannot exclude the possibility that there are other potential mechanisms. For instance, YouTubers might not only avoid competition to other YouTubers and acquire a more stable audience when they upload less popular content, but the characteristics of their viewers may change, too. Viewers of less popular content could be generally less ad averse or have a higher valuation of the video content such that they are willing to endure more ads.

Second, I cannot evaluate the effect of advertising on welfare, because I lack measures for consumer and producer surplus. Although I demonstrate that advertising leads to more content differentiation – which is likely to raise consumer surplus (Brynjolfsson et al., 2003) – the viewers must also pay an increased ad “price”, which works into the opposite direction. Since I obtain no estimates for the viewers' ad aversion, my setup does not answer which effect overweights. On the producer side, I remain agnostic about the effect of advertising on the surplus of YouTube itself, the YouTubers, and the advertisers. YouTube as a platform is likely to benefit from advertising, though. Advertising leads to more content differentiation, which attracts more viewers; more viewers, in turn, generate more ad revenue. Likewise, the YouTubers' surplus benefits from an increase in ad revenue; it is, however, unclear how their utility from covering different topics than before is affected. Finally, the advertisers' surplus may go up or down. On the one hand, a higher ad quantity makes it more likely that potential customers click on their ads and buy their products. On the other hand, the advertisers cannot influence where exactly their ads appear, whereby it is unclear how well the audience is targeted. Hence, it is possible that the additional costs of advertising surmount the additional revenues.

The third limitation of my paper is that the YouTubers' per-view-revenue from advertising is unaffected by the degree of targeting. Media outlets' revenue per ad usually increases in the degree of targeting, because the advertisers' willingness to pay is higher. On YouTube, in contrast, the price per ad is constant, whereby my results cannot be extrapolated to an environment where the per-ad-revenue increases if a narrow and specific audience is attracted. It is likely, however, that the effect of an increase in the feasible number of ad breaks was higher, because the YouTubers had an additional incentive to differentiate their content.

Finally, I do not discuss any concerns related to commercial media bias, i.e., advertisers exerting pressure on the media outlets' content decisions. As argued, however, there is no direct relationship between YouTubers and advertisers whose ads appear as breaks during the videos, so the issue is of small importance in my application. Yet, it is possible that commercial media bias arises from product placement contracts between advertisers and YouTubers, for instance, if the advertisers want their products to appear within friendly and uncontroversial videos; studying the relationship between product placement and commercial media bias on YouTube would be an interesting question for further research.

A. Robustness checks

This section probes the robustness of my results. In particular, I show that the main results from Section 6.1 are robust to using an alternative observation period, to an alternative selection of YouTubers, to an alternative classification of the treatment group, and to alternative definitions of the instrument $close_i$ and of the dependent variable $Popular_{vit}$. In addition to that, I report the results of placebo regressions that support the plausibility of the exclusion restriction, I conduct several robustness checks on the results from Section 7, and I probe the validity of the empirical strategy when studying video quality.

A.1. Alternative observation period

First, I show that the results from Section 6.1 are robust to using an alternative observation period. As argued in Section 4.3, I cannot extend the analysis to earlier or later points in time; I can, however, select a shorter observation period. Table 18 shows the results from estimating equations (2) and (3) on observations from Jan 2014 to July 2016 only (hence, I exclude twelve months before, and six months after Oct 2015). While the potentially biased OLS estimates in columns 1 to 3 are close to their counterparts in Table 3, the 2SLS estimates in columns 4 to 6 and the reduced form estimates in columns 7 to 9 are smaller by a third. This is no surprise: the event study in Section 6.2 illustrates that the effect of an increase in the feasible number of ad breaks on the probability to upload popular content becomes stronger over time. Thus, excluding the last six months from the analysis results in smaller estimates.

A.2. Alternative selections of YouTubers

Next, I demonstrate that the results from Section 6.1 are robust to alternative selections of YouTubers. As argued in Section 4.3, the final dataset includes only YouTubers whose median video duration before Oct 2015 is smaller than 10, because I want to focus on YouTubers who were ignorant of the ten minutes trick before the launch of the new ad break tool. One could argue, however, that the selection is too loose. For instance, if a YouTuber's median video duration before Oct 2015 is equal to 9, a large share of her videos may already be ten minutes or longer, so she could have come across the ten minutes trick before the new ad break tool was launched. To rule out concerns about the selection of YouTubers, I estimate regression equations (2) and (3) on two subsamples: first, a subsample of YouTubers whose median video duration before Oct 2015 is smaller than 7.5, second, a subsample of YouTubers whose 90^{th} percentile of the distribution of video durations (not the median) is smaller than 10.

Tables 19 and 20 show the results. The potentially biased OLS estimates in columns 1 to 3 resemble their counterparts in Table 3. Similarly, the 2SLS estimates in columns 4 to 6 are close to the estimates based on the entire dataset. The first stage as well as the reduced form estimates (columns 7 to 9), however, are nearly twice as large as their counterparts in Table 3. A potential explanation is that the average YouTuber whom I consider in this section has more scope to react to the launch of the new ad break tool than the average YouTuber from the main analysis, which matches the considerations from Section 5.2.3. In sum, I do not find evidence of my main results being sensitive to alternative selections of YouTubers.

A.3. Alternative classifications of the treatment group

This section shows that the results from Section 6.1 are robust to alternative classifications of the treatment group. In particular, I show that neither the five percentage point cutoff nor considering only a YouTuber's videos between ten and fourteen minutes drive my results.

Table 21 shows the 2SLS estimates from using two alternative cutoffs; YouTubers are classified as treated if their share of videos between ten and fourteen minutes has increased by at least one (columns 1 to 3) or by at least ten percentage points (columns 4 to 6). While the estimates in columns 1 to 3 are close to their counterparts in Table 3, the estimates in columns 4 to 6 are larger by a third: the probability to upload popular content decreases by around 35 percentage points for YouTubers in the treatment relative to the control group. The result is plausible: the average effect of an increase in the feasible number of ad breaks on the probability to upload popular content is stronger for YouTubers who increase their share of videos that are ten minutes or longer to a higher extent.

Next, I classify a YouTuber as treated if she increased her share of videos that are ten minutes or longer (instead of ten to fourteen minutes) by at least five percentage points. Table 22 shows the results. The potentially biased OLS estimates in columns 1 to 3 are close to zero and not statistically significant. The 2SLS estimates in columns 4 to 6 are negative and statistically significant at the 1%-level, but smaller in absolute value than their counterparts in Table 3. A potential explanation is that considering *all* videos that are ten minutes or longer leads to more noise in the estimation, for instance, because videos that are more than "just" longer than ten minutes are less likely to indicate that a YouTuber exploits the ten minutes trick. Finally, the reduced form estimates in columns 7 to 9 are similar to the results from Section 6.1.

A.4. Alternative definitions of the instrument

Next, I confirm that the results from Section 6.1 are robust to alternative definitions of the instrument $close_i$: While it is equal to a YouTuber's *median* video duration before Oct 2015 in the main analysis, $close_i$ corresponds to the 75^{th} and to the 90^{th} percentile of the distribution of her video durations here. These two alternative definitions of $close_i$ may better capture a YouTuber's "closeness" to the ten minutes threshold before Oct 2015.

Table 23 shows the results from a 2SLS estimation of equations (2) and (3) using a YouTuber's 75^{th} percentile (columns 1 to 3), and using a YouTuber's 90^{th} percentile of the distribution of video durations before Oct 2015 (columns 4 to 6) as an instrument for D_i. The estimates in columns 1 to 3 are negative, but smaller in absolute value than their counterparts in Table 3; they are also less statistically significant. The estimates in columns 4 to 6, in contrast, are larger than their counterparts in Table 3. The first stage estimates and the first F-statistics, however, are in both cases much smaller than in Table 3. Hence, the 75^{th} and the 90^{th} percentiles of the distribution of a YouTuber's video durations before Oct 2015 have less power to predict a YouTuber's treatment status D_i than the median.

A.5. Alternative definitions of popular content

Here, I show that the results from Section 6.1 are robust to alternative definitions of popular content. To this end, I generate four alternative measures. First, I assign a dummy equal to one to all videos that are given a keyword from the upper half percent, second, I assign a dummy equal to one to all videos that are given a keyword from the upper two percent of the distribution of most-viewed keywords (see Section 4.2). Third, instead of using a share, I classify a fixed number of keywords per month per category as popular – 250 keywords for the categories "Entertainment", "People & Blogs", and "Let's Play", where I have the most observations, and 100 keywords for the remaining categories – and assign a dummy equal to one to all videos given popular a keyword such defined. Finally, instead of considering the views, for each month, for each category, I compute how many *Likes* a certain keyword has attracted and rank them in descending order; the upper one percent of this distribution is then classified as popular and all videos given such a keyword are assigned a dummy equal to one. Table 26 provides an overview of how these measures are correlated.

Table 27 shows the results from a 2SLS estimation of equations (2) and (3) using the four alternative definitions of $Popular_{vit}$. In columns 1 to 3, the estimates for β are negative, but much smaller in absolute value than their counterparts in Table 3 and not statistically significant. The estimates in columns 4 to 6, in contrast, are larger by a third than the estimates in Table 3 and statistically significant at the 1%-level. When I use a fixed number of keywords per category per month to define popular content (columns 7 to 9), the estimates are negative and statistically significant at the 1%-level, but around a fourth smaller than in Table 3. Finally, when I consider the keywords' number of Likes instead of their views in columns 10 to 12, the estimates are close to their counterparts in Table 3.

A.6. Placebo regressions

In this section, I conduct a series of placebo regressions to support the plausibility of the exclusion restriction as discussed in Section 5.2.3. To this end, I augment the reduced form equation (4) to

$$Popular_{vit} = \gamma^{Placebo} close_i * fakepost_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + v_{vit} | \, t \leq 33, \qquad (16)$$

where in the first placebo regression, $fakepost_t$ is equal to one if $t \geq 3$, in the second placebo regression $fakepost_t$ is equal to one if $t \geq 4$, and so on; I run 29 placebo regressions in sum. If $close_i$ has no direct effect on $Popular_{vit}$, all estimates for $\gamma^{Placebo}$ should be close to zero and not statistically significant. The idea is similar to the event study in Section 5.2.3: YouTubers with different values of $close_i$ must not have been on different trends in terms of $Popular_{vit}$ before Oct 2015.

Of 29 placebo regressions, the estimate for $\gamma^{Placebo}$ is in three cases statistically significant at the 5%-level; these estimates are, however, positive. Thus, the results provide additional support for the plausibility of the exclusion restriction.

A.7. Robustness checks: mechanism

Next, I conduct several robustness checks on the results from Section 7. Since I know from Section 7.1 that the dependent variables $Popular_{vit}$ and $Competitive_{vit}$ are highly correlated, I do not repeat all the analyses from above, though. Instead, I focus on the robustness checks on the dependent variable as such, i.e., I study alternative definitions of competitive content and I run a series of placebo regressions to support the plausibility of the exclusion restriction from Section 7.2. In addition, I provide robustness checks on the commentator analysis in Section 7.4.

A.7.1. Alternative definitions of competitive content

Analogous to Appendix A.5, I show that the results from Section 7.3.1 are robust to alternative definitions of competitive content. Here, I generate three alternative measures. First, I assign a dummy equal to one to all videos that are given a keyword from the upper half percent, second, I assign a dummy equal to one to all videos that are given a keyword from the upper two percent of the distribution of most-*used* keywords (see Section 7.1). Third, instead of using a share, I classify a fixed number of keywords per month per category as competitive – 250 keywords for the categories "Entertainment", "People & Blogs", and "Let's Play", where I have the most observations, and 100 keywords for the remaining categories – and assign a dummy equal to one to all videos given competitive a keyword such defined. Table 28 provides an overview of how these measures are correlated.

Table 29 shows the results from a 2SLS regression of equations (2) and (3) using the three alternative definitions of $Competitive_{vit}$. All estimates are negative and statistically significant at the 1%-level. In columns 1 to 3, the estimates for β are similar to their counterparts in Table 3; the estimates in columns 4 to 9, in contrast, are a fourth to a fifth smaller in absolute value.

A.7.2. Placebo regressions

Analogous to Appendix A.6, I conduct a series of placebo regressions to support the plausibility of the exclusion restriction as discussed in Section 7.3.2. To this end, I augment the reduced form equation (9) to

$$Competitive_{vit} = \gamma^{Placebo'} close_i * fakepost_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + v_{vit} | \; t \leq 33, \quad (17)$$

where in the first placebo regression, $fakepost_t$ is equal to one if $t \geq 3$, in the second placebo regression $fakepost_t$ is equal to one if $t \geq 4$, and so on; as above, I run 29 placebo regressions in sum. Of 29 placebo regressions, the estimate for $\gamma^{Placebo'}$ is in four cases statistically significant at the 5%-level; these estimates are, however, positive. Thus, the results provide additional support for the plausibility of the exclusion restriction.

A.7.3. Alternative cutoffs in the commentator analysis

Here, I show that the results from Section 7.4 are robust to alternative comment cutoffs. To this end, I restrict the analysis to all advertising YouTubers who received (i) more than one hundred, (ii) more than fifty, (iii) more than ten, and (iv) at least one comment before and after Oct 2015.

The results in Table 24 confirm that the measure $fluctuation_i$ may lead to unreasonable results when the total number of comments is small. When I restrict the analysis to YouTubers with at least one hundred comments (columns 1 to 3) or to YouTubers with at least 50 comments before and after Oct 2015 columns (4 to 6), the potentially biased OLS estimate is smaller, while the 2SLS and the reduced form estimates are larger than their counterparts in Table 14. In contrast to that, when I restrict the analysis to YouTubers with at least ten comments before and after Oct 2015 (columns 7 to 9), the OLS is estimate larger, and the 2SLS and the reduced form estimates are smaller than in the main part and not statistically significant. Finally, when I consider all YouTubers who have received at least one comment (columns 10 to 12), the 2SLS and the reduced form estimate even switch their sign and become positive, but are not statistically significant.

A.8. Validity checks: Quality

Finally, I check if the empirical strategy from Section 5 is valid when I use $\frac{Likes}{Likes+Dislikes}_{vit}$ and $log(Views)_{vit}$ as dependent variables.

A.8.1. Exclusion restriction

To confirm the plausibility of the exclusion restriction, I conduct two further event studies. In particular, I estimate the augmented reduced form regression equations

$$\frac{Likes}{Likes + Dislikes}_{vit} = \sum_{t=1}^{33} \gamma_t'' close_i * pre_t + \sum_{t=35}^{49} \gamma_t'' close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + v_{vit}. \quad (18)$$

and

$$log(Views)_{vit} = \sum_{t=1}^{33} \gamma_t''' close_i * pre_t + \sum_{t=35}^{49} \gamma_t''' close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + v_{vit}. \quad (19)$$

by OLS. If $close_i$ has no impact on the dependent variables, then all estimates for γ_t'' and γ_t''', $t \in [1, 33]$, should be close to zero without being statistically significant.

Figures 21 and 22 show the results. The estimates for γ_t'' and γ_t''', $t \in [1, 33]$ are not statistically significant and fluctuate around zero, which supports the plausibility of the exclusion restriction. Yet, the lion's share of the estimates is not statistically significant at the 5%-level after Oct 2015, either.

A.8.2. Video duration, likes, and views

Next, I check if video duration as such affects the dependent variables $\frac{Likes}{Likes+Dislikes}_{vit}$ and $log(Views)_{vit}$ (see Section 5.2.4). To this end, I estimate

$$\frac{Likes}{Likes + Dislikes}_{vit} = \delta'' duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \epsilon_{vit} \mid t \leq 33 \quad (20)$$

and

$$log(Views)_{vit} = \delta''' duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \epsilon_{vit} \mid t \leq 33 \quad (21)$$

by OLS.

Table 5 shows the results. The size of the estimates for δ'' (columns 1 to 3), though statistically significant at the 1%-level, is negligible: a one second increase in video duration corresponds to a 0.0001 percentage point increase in the fraction of likes. The estimates for δ''' in columns 4 to 6, though, are relatively large and statistically significant at the 1%-level, too. According to these estimates, one further second in video duration leads on average to about 1.5 percent more views. These estimates may reflect the algorithmic drift discussed in Section 9.2. YouTube wants to keep its viewers as long as possible on the platform to show as many ads as possible to them. As a result, longer videos get higher rankings and are watched more often.

B. Further discussions

This section revisits a number of topics that could not be covered in the main part of the paper. In particular, I discuss the consequences of misclassifying advertising and non-advertising YouTubers, I show that no YouTube platform event beyond the launch of the new ad break tool affects my results, and I discard a YouTuber learning effect as a potential economic mechanism behind content differentiation.

B.1. Misclassification of advertising and non-advertising YouTubers

As explained in Section 4.1, I cannot retrieve data on the YouTubers' monetization settings on the video level. Instead, I pick twenty randomly drawn videos per YouTuber, and classify her as advertising YouTuber if I detect at least one ad break. In this section, I amplify how measurement errors during this procedure could affect my results. In addition, I discuss the consequences of sample migration between advertising and non-advertising YouTubers.

B.1.1. Potential consequences of measurement error

In this section, I illustrate that a potential measurement error would have only minor consequences. First, note that I could erroneously classify an advertising YouTuber as non-advertising, but not vice versa: if a YouTuber never permits for ad breaks, my algorithm cannot classify her as "advertising" by definition. Second, note that I do not use the classification dummy in a regression framework; hence, the regression results do not suffer from an errors-in-variables bias (e.g., Durbin, 1954). Yet, I *split* my sample into advertising and non-advertising YouTubers. Thus, misclassifying some advertising as non-advertising YouTubers might lead to selection bias in the subsamples.

If I misclassified some advertising as non-advertising YouTubers, the estimates in Table 3 may be too large. YouTubers who fall through the grid of the algorithm seldom permit for ad breaks and do not follow strict commercial incentives. Thus, they are on average more reluctant to adapt their content after Oct 2015 than the average YouTuber whom the algorithm detects. On the other hand, the YouTubers whom I missed might not even increase their share of videos between ten minutes and fourteen minutes. Thus, they are not affected by the instrument $close_i$ and their first stage is equal to zero. In this case, the LATE (see Section 5.2.2) was the same whether or not I classified some advertising as non-advertising YouTubers.

If some advertising YouTubers were included into the subsample of non-advertising YouTubers, the estimates in Table 6 may be too large, too. This would, however, strengthen my results: Section 6.2.3 demonstrates that there is no effect of an increase in the feasible number of ad breaks on the non-advertising YouTubers' content choice; if the estimates were even closer to zero, the validity check would be even more convincing.

B.1.2. Potential consequences of sample migration

An advertising YouTuber may have been non-advertising in the past and vice versa. Potential sample migration between advertising and non-advertising YouTubers, however, is unproblematic for three reasons. First, I do not directly compare advertising to non-advertising YouTubers. Second, many advertising YouTubers may have started as non-advertising YouTubers in the beginning of their career. If they became advertising YouTubers as a result of the treatment, they may have adapted their content with a delay, which may lead to an underestimation of the effect of advertising on content differentiation. Finally, if former advertising YouTubers have migrated to the subsample of non-advertising YouTubers, I might overestimate the main effect , which would – as argued in the previous subsection – make the validity check more convincing.

B.2. Platform events during the observation period

Next, I provide a systematic review of all platform "events" during my observation period, i.e., technical novelties or changes in YouTube's monetization policy beyond the launch of the new ad break tool. Note that an event can only affect my results if it is correlated to a YouTuber's probability to upload popular content *and* to her value of $close_i$ – no such event exists during the observation period. Since YouTube has no serious competitors, I remain agnostic about events at competing video sharing platforms.

B.2.1. Data collection

I collect information on all events from the YouTube Creators Blog, which announces YouTube news, introduces technical features, and gives general advice to YouTubers.[32] In a first step, I retrieve all blog posts from Jan 2013 to Jan 2017. Next, I manually exclude any post that does not deal with a platform event, such as YouTube promotion for academies, awards, (real world) events, and YouTuber portraits. The remaining 42 posts are listed in Table 30. In a last step, I review all posts from Table 30 and indicate if a YouTuber's monetization options or her probability to upload popular content could be affected. Thirteen events require further investigation; I discuss them chronologically.

B.2.2. Platform events in 2013

First, in March, YouTubers' access to their financial data changed. This event applies to all YouTubers equivalently, has no effect on their content choice, and is therefore unproblematic.

In May, selected YouTubers from the U.S., and in October, selected YouTubers worldwide were given the option to raise a subscription fee of 0.99$ per month. The pilot was,

[32]See `youtube-creators.googleblog.com/` (May 2019).

however, extremely limited: not even 100 YouTubers worldwide participated.[33] Thus, my results are unlikely to be affected by these events.

Next, YouTube launched its "Fan Finder": a YouTuber could let the platform turn one of her videos into an "ad" and show it to viewers of a different channel in place of a conventional ad; this was supposed to enlarge a YouTuber's fan base. Since YouTubers were asked to produce special videos that advertise their channel, the event may have affected their content choice. Yet, all YouTubers with at least $1,000$ subscribers could participate and there were no restrictions on the advertising video's duration. Hence, the event is not correlated to $close_i$ and thereby unproblematic.

Finally, live streams became technically feasible in December and may have influenced YouTuber's content choice. The feature is open to all YouTubers, though. Hence, the event is not correlated to $close_i$ and cannot affect my results.

B.2.3. Platform events in 2015

In March, 360 degree videos became technically feasible. Similar to the live streams, the event may have influenced YouTubers' content choice, but since it is open to all YouTubers, there is no correlation to $close_i$.

YouTube Red, a paid subscription service that provides advertising-free streaming of all videos and exclusive original content was launched in October. The availability of YouTube Red is, however, limited to the US. Since my dataset includes only German YouTube channels, the event cannot affect my results.

In November, several virtual reality tools became available. Again, YouTubers' content choice may have been affected, but since the features are open to all YouTubers, there is no correlation to $close_i$.

B.2.4. Platform events in 2016

In January, YouTube launched a "Donate Button": users who click on the button can donate to a YouTuber after watching her video. As with the technical novelties from above, this may have influenced YouTubers' content choice. In addition, their monetization options were affected. Still, the feature is open to all YouTubers and thereby not correlated to $close_i$.

Next, in April, YouTube announced that it would withhold (not block) all ad revenue generated during copyright disputes. This event applies to all YouTubers equivalently, has no effect on their content choice, and is therefore unproblematic.

Mobile live streams became technically feasible in June, i.e., YouTubers could stream from their mobile devices. Similar to the "stationary" live streams from 2013, the event is not correlated to $close_i$ and cannot affect my results.

[33] E.g., www.fastcompany.com/3020553/the-most-popular-youtube-channels-might-start-charging-you-to-watch, www.bbc.com/news/business-22474715, or searchenginewatch.com/sew/news/2267170/youtube-launches-paid-channels-subscription-fees-start-at-usd099-per-month (May 2019).

In October, YouTube launched an optional feature for paid promotion disclosure: by checking the "video contains paid promotion" box in their settings, YouTubers can inform their audience about paid product placement and endorsements by third parties. This may influence their videos' content, but is unrelated to $close_i$.

Finally, in October, video end screens, that allow YouTubers to promote up to four different videos or playlists, became technically available. Although the event may have affected the YouTubers' content choice, the feature is open to all YouTubers, thereby not correlated to $close_i$, and hence unproblematic.

B.3. YouTuber learning effect

Here, I discuss a YouTuber learning effect as an alternative explanation for the results from Section 6: YouTubers copy the most popular content in the beginning of their career, but deviate from the mainstream when they become more experienced and start to develop a personal style. If such a learning effect was positively correlated with $close_i$, it could be the driving force behind the decrease in the probability to upload popular content after Oct 2015 rather than an increase in the feasible number of ad breaks per video.

Three arguments, however, speak against a YouTuber learning effect. First, there exists no plausible reason why YouTubers with a high value of $close_i$ would experience a stronger learning effect than YouTubers whose value of $close_i$ is low. See Section 5.2.3 for a detailed discussion on the independence of $close_i$.

Second, t_i controls for a YouTuber's average change in the probability to upload popular (or competitive) content over time. Columns 1 and 4 in Table 31 replicate the 2SLS results from Tables 3 and 11 and illustrate that a linear YouTuber learning effect is of minor importance. On the one hand, the estimates for β and β' are nearly unaffected when I control for t_i. On the other hand, the estimates for t_i, though negative, are extremely small. A YouTuber's probability to upload popular content decreases by 0.00008 percentage points for each additional video; similarly, her probability to upload competitive content decreases by 0.0003 percentage points for each additional video.

Third, allowing for a more flexible YouTuber specific time trend by adding t_i^2 and t_i^3 does not affect the estimates for β and β', either (columns 2, 3, 5, and 6 of Table 31). It becomes, however, obvious that the YouTuber specific time trend is not linear. For instance, columns 2 and 5 illustrate that a YouTuber's probability to upload popular or competitive content increases in the beginning, but decreases from around her 160^{th} video, which is consistent with the story from above. Note that the average number of videos per YouTuber is 99.3 and the median number of videos is 64. Thus, many YouTubers in my sample do not reach the turning point of 160. In sum, even though I find some evidence for a YouTuber learning effect, it is not the driving force behind the main results from Section 6.

References

Anderson, Ashton, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec (2013) "Steering User Behavior with Badges," in *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, 95–106: ACM.

Anderson, Chris (2006) *The long tail: Why the future of business is selling less of more*: Hachette Books.

Anderson, Michael and Jeremy Magruder (2012) "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *Economic Journal*, **122** (563), 957–989.

Anderson, Simon P., Andre De Palma, and Jacques-Francois Thisse (1992) *Discrete choice theory of product differentiation*: MIT press.

Anderson, Simon P. and Bruno Jullien (2016) "The advertising-financed business model in two-sided media markets," in *Handbook of Media Economics*, **1**: Elsevier, 41–90.

Anderson, Simon P. and Joel Waldfogel (2016) "Preference externalities in media markets," in *Handbook of Media Economics*, **1**: Elsevier, 3–40.

Angrist, Joshua D and Guido W Imbens (1995) "Identification and Estimation of Local Average Treatment Effects," Working Paper 118, National Bureau of Economic Research.

Aral, Sinan and Dylan Walker (2012) "Identifying Influential and Susceptible Members of Social Networks," *Science*, **337** (6092), 337–341.

Armstrong, Mark and Helen Weeds (2007) "Programme quality in subscription and advertising-funded television," *Unpublished working paper*.

Beebe, Jack H. (1977) "Institutional structure and program choices in television markets," *The Quarterly Journal of Economics*, **91** (1), 15–37.

Behringer, Stefan and Lapo Filistrucchi (2015) "Hotelling competition and political differentiation with more than two newspapers," *Information Economics and Policy*, **30**, 36–49.

Bond, Robert, Christopher Fariss, Jason J. Jones, Adamd. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler (2012) "A 61-million-person experiment in social influence and political mobilization," *Nature*, **489** (7415), 295–298.

Bourreau, Marc (2003) "Mimicking vs. counter-programming strategies for television programs," *Information Economics and Policy*, **15** (1), 35–54.

Brynjolfsson, Erik, Yu Hu, and Duncan Simester (2011) "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales," *Management Science*, **57** (8), 1373–1386.

Brynjolfsson, Erik, Yu Hu, and Michael D Smith (2003) "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers," *Management Science*, **49** (11), 1580–1596.

Bundesverfassungsgericht (2007) "Urteil des Ersten Senats vom 11. September 2007," 1 BvR 2270/05 - Rn. (1-213).

Chevalier, Judith and Dina Mayzlin (2006) "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, **43** (3), 345–354.

Coase, Ronald Harry (1974) "The market for goods and the market for ideas," *The American Economic Review*, **64** (2), 384–391.

d'Aspremont, Claude, J Jaskold Gabszewicz, and J-F Thisse (1979) "On Hotelling's" Stability in competition"," *Econometrica*, 1145–1150.

De Palma, Andre, Victor Ginsburgh, Yorgo Y. Papageorgiou, and Jaques-Francois Thisse (1985) "The principle of minimum differentiation holds under sufficient heterogeneity," *Econometrica: Journal of the Econometric Society*, 767–781.

Downs, Anthony (1957) "An economic theory of political action in a democracy," *Journal of political economy*, **65** (2), 135–150.

Dukes, Anthony (2004) "The adverstising market in a product oligopoly," *The Journal of Industrial Economics*, **52** (3), 327–348.

Durbin, James (1954) "Errors in variables," *Revue de l'institut International de Statistique*, 23–32.

Easley, David and Arpita Ghosh (2013) "Incentives, gamification, and game theory: An economic approach to badge design," *ACM Transactions on Economics and Computation (TEAC)*, **4** (3), 16.

Economides, Nicholas (1986) "Minimal and maximal product differentiation in Hotellings duopoly," *Economics Letters*, **21** (1), 67–71.

Gabszewicz, Jean J., Didier Laussel, and Nathalie Sonnac (2001) "Press advertising and the ascent of the pensée unique," *European Economic Review*, **45** (4-6), 641–651.

——— (2002) "Press advertising and the political differentiation of newspapers," *Journal of Public Economic Theory*, **4** (3), 317–334.

——— (2004) "Programming and advertising competition in the broadcasting industry," *Journal of Economics & Management Strategy*, **13** (4), 657–669.

Gabszewicz, Jean and Jaques-Francois Thisse (1986) “Spatial competition and the location of firms,” in *Fundamentals of pure and applied economics*, **5**: Routledge, 1–71.

Gal-Or, Esther and Anthony Dukes (2003) “Minimum differentiation in commercial media markets,” *Journal of Economics & Management Strategy*, **12** (3), 291–325.

Garcia Pires, Armando (2014) “Media diversity, advertising, and adaptation of news to readers political preferences,” *Information Economics and Policy*, **28**, 28–38.

Gentzkow, Matthew, Jesse M Shapiro, and Daniel F Stone (2016) “Media bias in the marketplace: Theory,” in *Handbook of media economics*, **1**: Elsevier, 623–645.

George, Lisa (2007) “Whats fit to print: The effect of ownership concentration on product variety in daily newspaper markets,” *Information Economics and Policy*, **19** (3-4), 285–303.

George, Lisa and Joel Waldfogel (2003) “Who affects whom in daily newspaper markets?” *Journal of Political Economy*, **111** (4), 765–784.

Goldfarb, Avi and Catherine Tucker (2019) “Digital economics,” *Journal of Economic Literature*, **57** (1), 3–43.

Graitson, Dominique (1982) “Spatial competition a la Hotelling: A selective survey,” *The Journal of Industrial Economics*, 11–25.

Hamilton, James (2004) *All the news that’s fit to sell: How the market transforms information into news*: Princeton University Press.

Herman, Edward S and Robert Waterman McChesney (1997) *The global media: The new missionaries of corporate capitalism*: Burns & Oates.

Hotelling, Harold (1929) “Stability in Competition,” *Economic Journal*, **39**, 41–57.

Huang, Jason, David Reiley, and Nick Riabov (2018) “Measuring Consumer Sensitivity to Audio Advertising: A Field Experiment on Pandora Internet Radio.”

Kleibergen, Frank and Richard Paap (2006) “Generalized reduced rank tests using the singular value decomposition,” *Journal of econometrics*, **133** (1), 97–126.

Lancaster, Kelvin (1990) “The economics of product variety: A survey,” *Marketing science*, **9** (3), 189–206.

Luca, Michael (2016a) “Reviews, reputation, and revenue: The case of Yelp. com,” Unpublished.

——— (2016b) “User-generated content and social media,” in *Handbook of Media Economics*, **1**: Elsevier, 563–592.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014) "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, **104** (8), 2421–2455.

McChesney, Robert D (2004) *The problem of the media: US communication politics in the twenty-first century*: NYU Press.

McCrary, Justin (2008) "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, **142** (2), 698–714.

Peitz, Martin and Tommaso M. Valletti (2008) "Content and advertising in the media: Pay-tv versus free-to-air," *international Journal of industrial organization*, **26** (4), 949–965.

Puglisi, Riccardo and James M Snyder (2016) "Empirical studies of media bias," in *Handbook of media economics*, **1**: Elsevier, 647–667.

Seamans, Robert and Feng Zhu (2014) "Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers," *Management Science*, **60** (2), 476–493.

Steiner, Peter O. (1952) "Program patterns and preferences, and the workability of competition in radio broadcasting," *The Quarterly Journal of Economics*, **66** (2), 194–223.

Stock, James H and Motohiro Yogo (2002) "Testing for weak instruments in linear IV regression."

Sun, Monic and Feng Zhu (2013) "Ad revenue and content commercialization: Evidence from blogs," *Management Science*, **59** (10), 2314–2331.

Waldfogel, Joel (2007) *The tyranny of the market: Why you can't always get what you want*: Harvard University Press.

——— (2017) "How digitization has created a golden age of music, movies, books, and television," *Journal of economic perspectives*, **31** (3), 195–214.

——— (2018) *Digital Renaissance: What Data and Economics Tell Us about the Future of Popular Culture*: Princeton University Press.

Wang, Zhongmin (2010) "Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews," *The BE Journal of Economic Analysis & Policy*, **10** (1), 1–35.

Weeds, Helen (2013) "Programme Quality in Subscription and Advertising-Funded Television."

Wilbur, Kenneth C (2008) "A two-sided, empirical model of television advertising and viewing markets," *Marketing science*, **27** (3), 356–378.

Zhang, Xiaoquan (Michael) and Feng Zhu (2011) “Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia,” *American Economic Review*, **101** (4), 1601–1615.

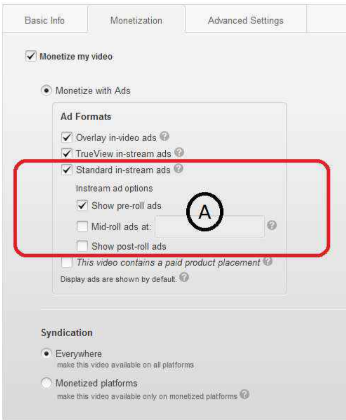


Figure 1: Old ad break tool (before Oct 2015).

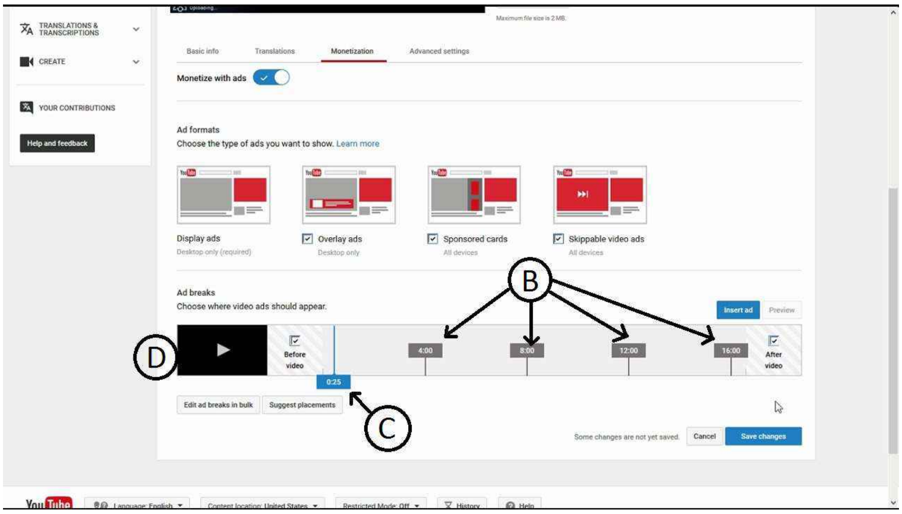


Figure 2: New ad break tool (after Oct 2015).

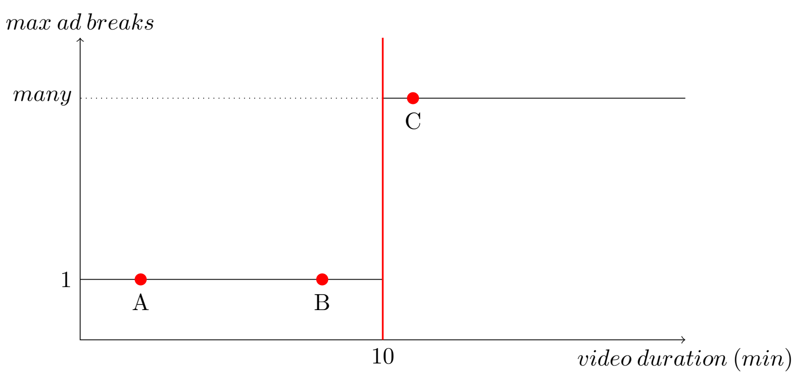


Figure 3: Stylized example of the identification strategy.

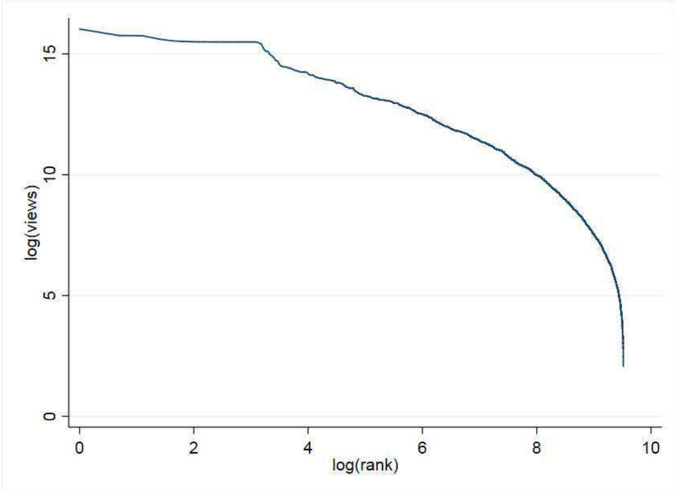


Figure 4: Log-log plot of the number of views a keyword attracts and its associated rank in the category "Science & Technology" in March 2015.

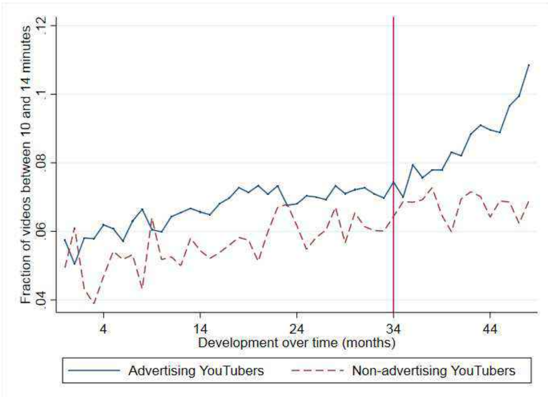


Figure 5: Trends in advertising vs. non-advertising YouTubers' share of videos between ten and fourteen minutes. The vertical line depicts Oct 2015.

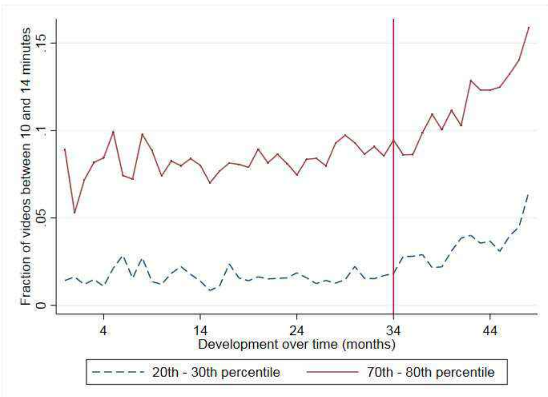
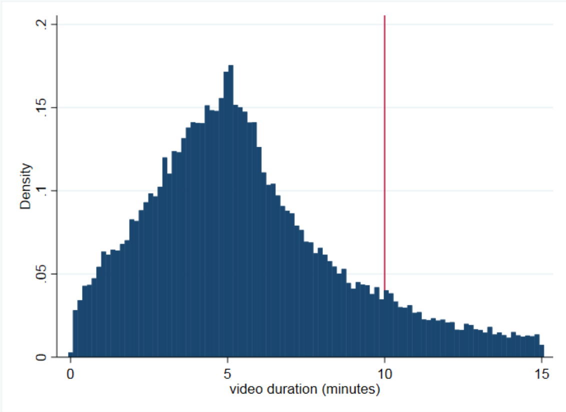


Figure 6: Trends in advertising YouTubers' share of videos between ten and fourteen minutes. The vertical line depicts Oct 2015.



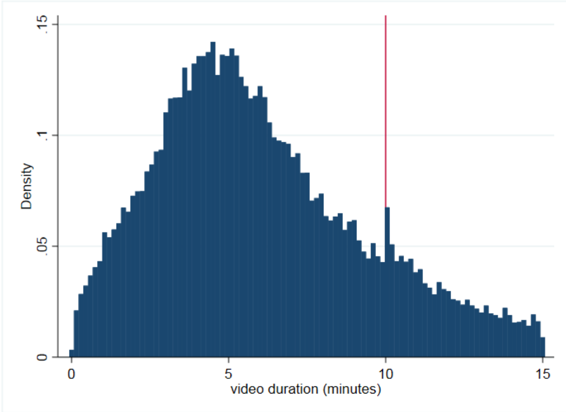
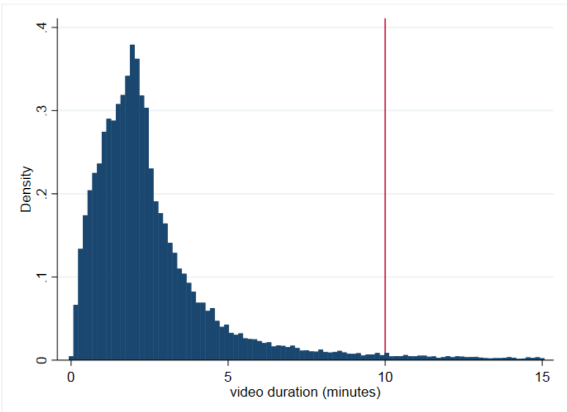


Figure 7: Histogram of the distribution of video durations *before* Oct 2015 for YouTubers from the 70th to 80th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

Figure 8: Histogram of the distribution of video durations *after* Oct 2015 for YouTubers from the 70th to 80th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.



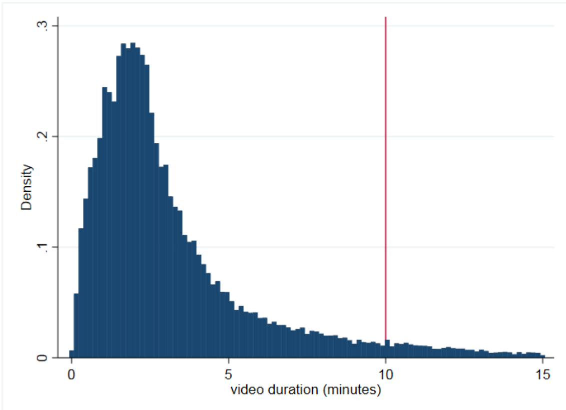


Figure 9: Histogram of the distribution of video durations *before* Oct 2015 for YouTubers from the 20th to 30th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

Figure 10: Histogram of the distribution of video durations *after* Oct 2015 for YouTubers from the 20th to 30th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

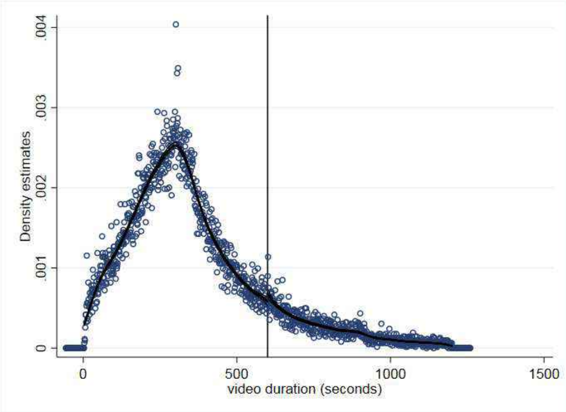


Figure 11: McCrary test of the distribution of video durations *before* Oct 2015 for YouTubers from the 70th to 80th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

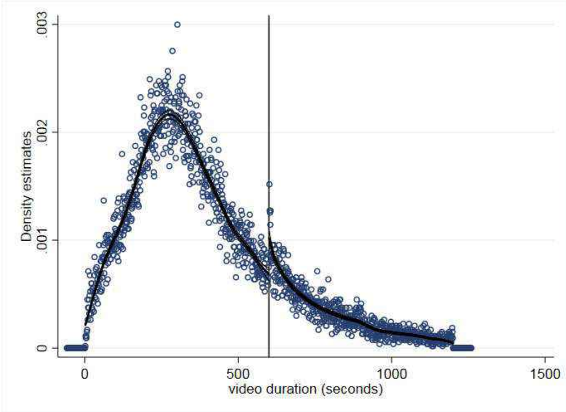


Figure 12: McCrary test of the distribution of video durations *after* Oct 2015 for YouTubers from the 70th to 80th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

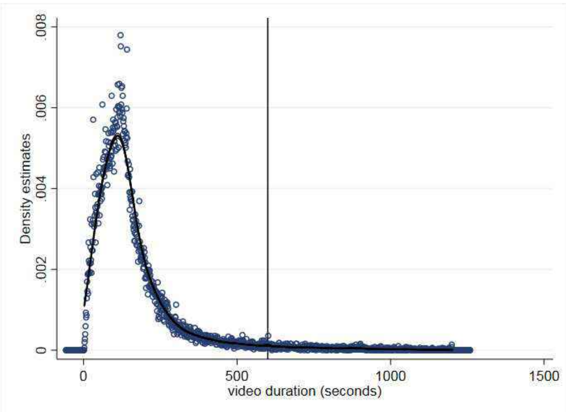


Figure 13: McCrary test of the distribution of video durations *before* Oct 2015 for YouTubers from the 20th to 30th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

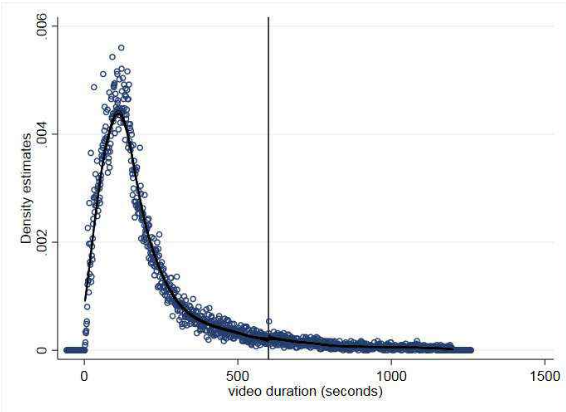


Figure 14: McCrary test of the distribution of video durations *after* Oct 2015 for YouTubers from the 20th to 30th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

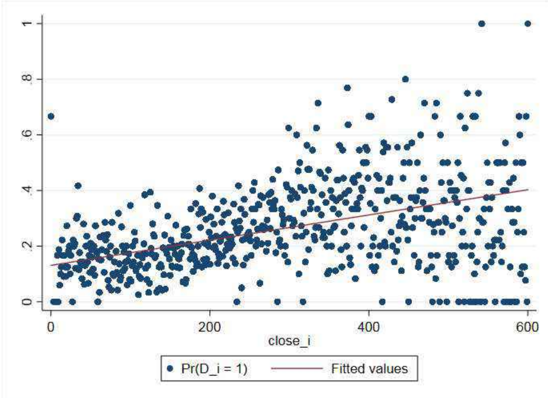


Figure 15: Plot of all values of $close_i$ on the associated average probability to be treated, $Pr(D_i = 1)$.

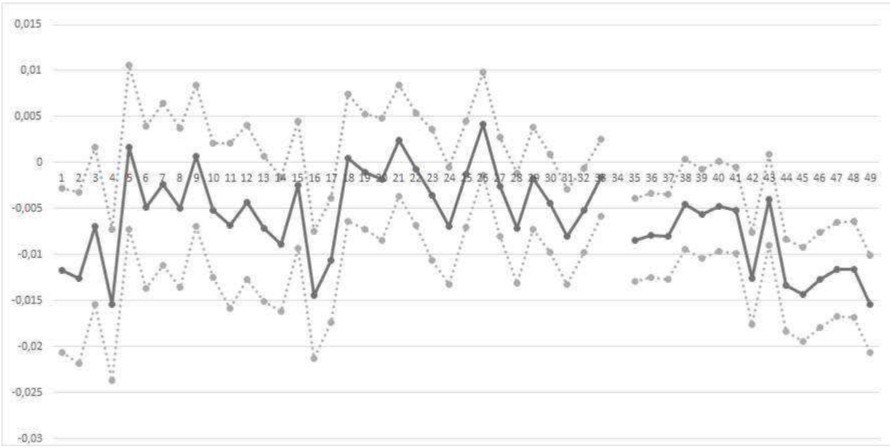


Figure 16: Event study popular content (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t$ from equation (5). The dashed line depicts a 95%-confidence interval.

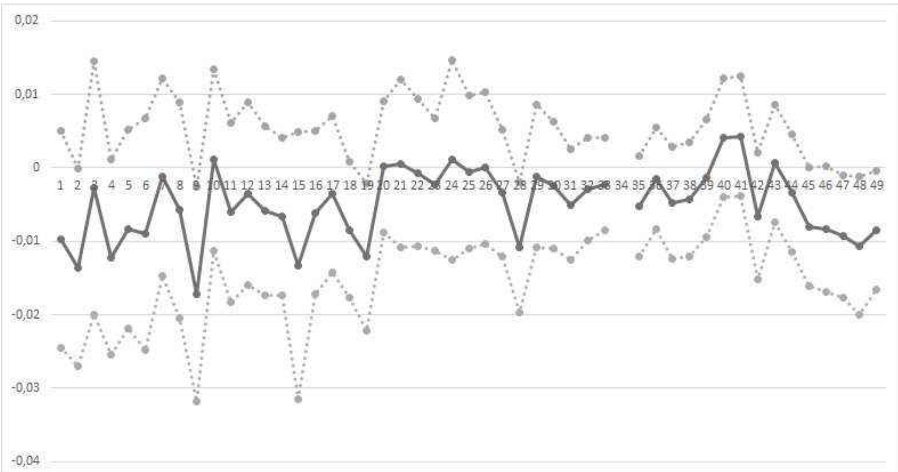


Figure 17: Event study popular content (non-advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t$ from equation (5). The dashed line depicts a 95%-confidence interval.

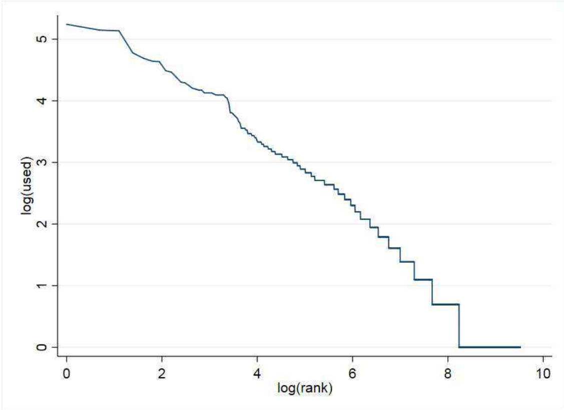


Figure 18: Log-log plot of the number of usages of a keyword and their associated rank in the category "Science & Technology" in March 2015.

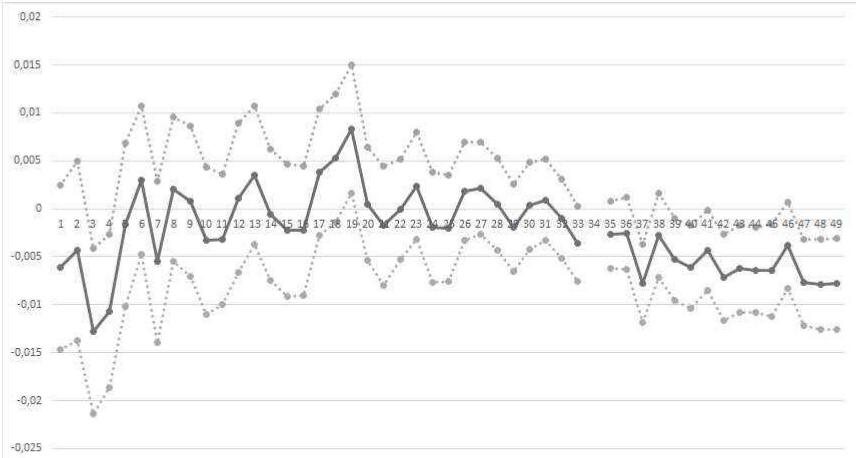


Figure 19: Event study competitive content (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t'$ from equation (9). The dashed line depicts a 95%-confidence interval.

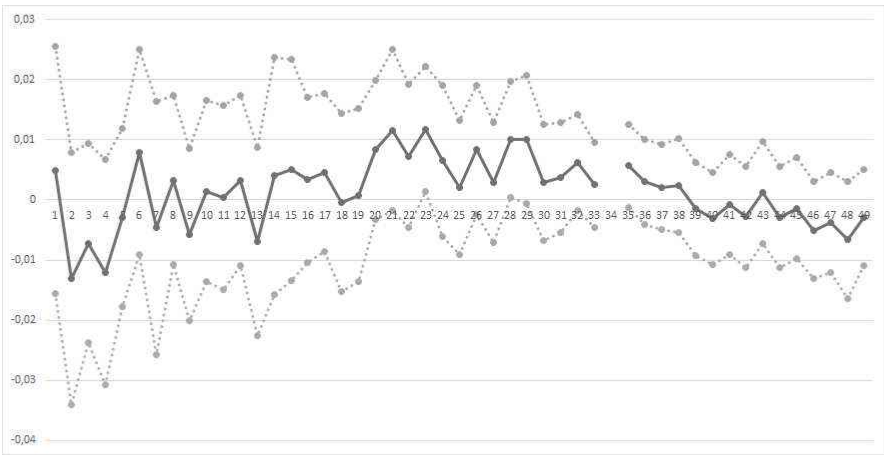


Figure 20: Event study competitive content (non-advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t'$ from equation (9). The dashed line depicts a 95%-confidence interval.

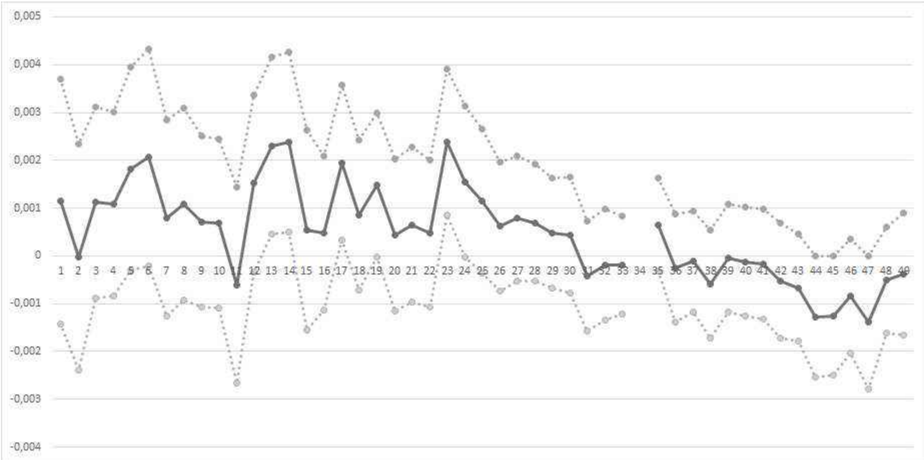


Figure 21: Event study likes/(likes+dislikes) (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t''$ from equation (18). The dashed line depicts a 95%-confidence interval.

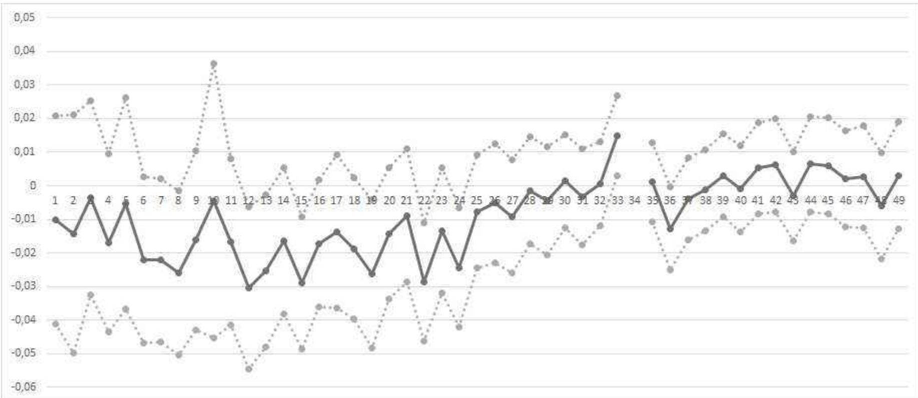


Figure 22: Event study log(views) (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t'''$ from equation (19). The dashed line depicts a 95%-confidence interval.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
$Popular_{vit}$	0.425	0.494	0	1	1,397,267
$Competitive_{vit}$	0.641	0.480	0	1	1,397,267
$Advertising_i$	0.764	0.425	0	1	15,877
$post_t$	0.475	0.499	0	1	1,397,267
D_i	0.226	0.418	0	1	15,877
$Duration_{vit}$	6.411	13.341	0	1440.033	1,397,267
$Subscribers_i$	18,234.506	138,282.229	0	6,581,640	15,877
$Film\&Animation_{vit}$	0.086	0.280	0	1	1,397,267
$Cars\&Vehicles_{vit}$	0.081	0.272	0	1	1,397,267
$Music_{vit}$	0.025	0.155	0	1	1,397,267
$Pets\&Animals_{vit}$	0.026	0.159	0	1	1,397,267
$Sports_{vit}$	0.085	0.278	0	1	1,397,267
$Travel\&Events_{vit}$	0.056	0.229	0	1	1,397,267
$Let'sPlay_{vit}$	0.085	0.278	0	1	1,397,267
$People\&Blogs_{vit}$	0.202	0.402	0	1	1,397,267
$Comedy_{vit}$	0.015	0.121	0	1	1,397,267
$Entertainment_{vit}$	0.201	0.401	0	1	1,397,267
$HowTo\&Style_{vit}$	0.064	0.245	0	1	1,397,267
$Education_{vit}$	0.046	0.210	0	1	1,397,267
$Science\&Technology_{vit}$	0.014	0.119	0	1	1,397,267
$Nonprofit\&Activism_{vit}$	0.015	0.120	0	1	1,397,267
$I(1^{st}to10^{th})_{vit}$	0.673	0.469	0	1	1,397,267
$I(10^{th}to25^{th})_{vit}$	0.581	0.493	0	1	1,397,267
$I(25^{th}to50^{th})_{vit}$	0.548	0.498	0	1	1,397,267
$I(50^{th}to75^{th})_{vit}$	0.390	0.488	0	1	1,397,267
$I(75^{th}to100^{th})_{vit}$	0.284	0.451	0	1	1,397,267
$SumAffiliations_{vit}$	2.472	1.160	0	5	1,397,267
$Likes_{vit}$	631.945	5,993.188	0	1,269,177	1,397,267
$Dislikes_{vit}$	33.452	532.98	0	149,614	1,397,267
$Views_{vit}$	35,098.614	564,233.77	0	337,832,408	1,397,267

Notes: This table presents the summary statistics of all variables used in the analysis. The variables $Popular_{vit}$, $Competitive_{vit}$, $Advertising_i$, $post_t$, D_i, all percentile indicators, and all category indicators are dummy variables. The variables $Advertising_i$, D_i, $close_i$, and $Subscribers_i$ are available only on the YouTuber level.

Table 2: McCrary test

Figure	Estimate
70th to 80th percentile pre Oct 2015	0.1654***
(Figure 11)	(0.0640)
70th to 80th percentile post Oct 2015	0.4049***
(Figure 12)	(0.0604)
20th to 30th percentile pre Oct 2015	0.0035
(Figure 13)	(0.1609)
20th to 30th percentile post Oct 2015	0.2659**
(Figure 14)	(0.1356)

Notes: Standard errors in parentheses. The estimates depict discontinuity estimates (log difference in height) of a McCrary test with bin width 1 and band width 60. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Main results

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red.Form (9)
$D_i * post_t$	0.009	0.006	0.006	-0.209***	-0.200***	-0.192***			
	(0.0078)	(0.0077)	(0.0077)	(0.0505)	(0.0491)	(0.0480)			
$close_i * post_t$							-0.006***	-0.006***	-0.006***
							(0.0013)	(0.001)	(0.0013)
First stage				0.0286***	0.0286***	0.0290***			
				(0.0023)	(0.0023)	(0.0024)			
F-test of excluded instruments				144.13	143.85	151.32			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Popular content – Evidence from the video level

	Popular (1)	Popular (2)	Popular (3)	Non-pop. (4)	Non-pop. (5)	Non-pop. (6)
$close_i * post_t$	0.0005	0.0006	0.0005	0.003**	0.003**	0.003**
	(0.0012)	(0.0011)	(0.0011)	(0.0012)	(0.0012)	(0.0011)
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	9,855	9,855	9,855	10,248	10,248	10,248
Videos	477,532	477,532	477,532	589,468	589,468	589,468

Notes: Robust standard errors in parentheses. The dependent variable is $I(\geq 10)_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is ten minutes or longer, and 0 otherwise. All estimates are obtained by OLS and based on using the advertising YouTubers only. In addition, the estimates in Columns 1 to 3 are based on videos classified as popular. The estimates in Columns 4 to 6 are based on videos classified as non-popular. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Video duration and popular content

	OLS (1)	OLS (2)	OLS (3)
video	0.0000760	0.0000125	0.0000119
duration	(0.000103)	(0.000101)	(0.000101)
Time FE	X	X	X
YouTuber FE	X	X	X
Controls		X	X
YouTuber Time Trend			X
YouTubers	10,113	10,113	10,113
Videos	566,079	566,079	566,079
R^2	0.404	0.412	0.412

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, the estimates are based on a regression that excludes all months $t \geq 34$. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Main results non-advertising YouTubers

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	0.033**	0.0364**	0.0364**	0.0227	-0.0121	-0.0125			
	(0.0146)	(0.0143)	(0.0142)	(0.0938)	(0.0894)	(0.0886)			
$close_i * post_t$							0.0005	-0.0002	-0.0003
							(0.0022)	(0.0021)	(0.0021)
First stage				0.0237***	0.0237***	0.0239***			
				(0.0035)	(0.0035)	(0.0035)			
F-test of excluded instruments				45.74	45.81	47.70			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278
Videos	329,725	329,725	329,725	329,725	329,725	329,725	329,725	329,725	329,725

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the non-advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Main regression, subscribers $\geq 1,000$

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	0.009	0.008	0.007	-0.259***	-0.238***	-0.230***			
	(0.0099)	(0.0097)	(0.0098)	(0.0689)	(0.0662)	(0.0650)			
$close_i * post_t$							-0.007***	-0.007***	-0.007***
							(0.002)	(0.002)	(0.002)
First stage				0.028***	0.028***	0.029***			
				(0.003)	(0.003)	(0.003)			
F-test of excluded instruments				83.08	83.23	86.28			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	5,182	5,182	5,182	5,182	5,182	5,182	5,182	5,182	5,182
Videos	677,590	677,590	677,590	677,590	677,590	677,590	677,590	677,590	677,590

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers with at least $1,000$ subscribers are included. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Main regression, subscribers $< 1,000$

	OLS	OLS	OLS	2SLS	2SLS	2SLS	Red. Form	Red. Form	Red. Form
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$D_i * post_t$	0.015	0.008	0.009	-0.050	-0.069	-0.070			
	(0.012)	(0.012)	(0.012)	(0.082)	(0.082)				
$close_i * post_t$							-0.001	-0.002	-0.002
							(0.002)	(0.002)	(0.002)
First stage				0.0247^{***}	0.0245^{***}	0.0248^{***}			
				(0.0035)	(0.0036)	(0.0035)			
F-test of excluded instruments				48.26	47.79	48.81			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	5,416	5,416	5,416	5,416	5,416	5,416	5,416	5,416	5,416
Videos	389,952	389,952	389,952	389,952	389,952	389,952	389,952	389,952	389,952

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers with fewer than $1,000$ subscribers are included. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Heterogeneity along video categories

	Film and Animation (1)	Cars and Vehicles (2)	Music (3)	Pets and Animals (4)	Sports (5)	Travel and Events (6)	Let's Play (7)	People and Blogs (8)	Comedy (9)	Enter-tainment (10)	How To and Style (11)	Edu-cation (12)	Science and Techn. (13)	Nonprofit and Activ. (14)
$D_i * post_t$	-0.922***	-0.0762	2.712	0.170	-0.105	0.127	-0.576	-0.259**	-0.548	-0.255**	-0.235	-0.100	-0.651	-0.227
	(0.3405)	(0.0953)	(6.113)	(0.179)	(0.134)	(0.139)	(0.773)	(0.109)	(0.936)	(0.122)	(0.188)	(0.258)	(0.761)	(0.197)
First stage	0.0220***	0.044***	-0.003	0.037**	0.035***	0.030***	-0.009	0.031***	0.020	0.028***	0.028***	0.026**	0.021	0.047
	(0.006)	(0.010)	(0.017)	(0.017)	(0.009)	(0.010)	(0.009)	(0.006)	(0.023)	(0.005)	(0.008)	(0.012)	(0.017)	(0.031)
F-test of excluded instruments	13.49	21.08	0.04	4.99	15.34	8.41	0.98	32.15	0.72	28.32	10.99	4.99	1.70	2.32
Time FE	X	X	X	X	X	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X	X	X	X	X	X
YouTuber Time trend	X	X	X	X	X	X	X	X	X	X	X	X	X	X
YouTubers	2,302	1,543	1,382	838	1,776	1,764	827	4,724	769	4,207	1,462	906	868	430
Videos	93,616	87,945	25,512	25,963	96,645	62,041	82,650	200097	15,831	224,739	77,593	43,446	13,951	11,622

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. Each column displays the results of a 2SLS estimation including only the observations from one particular video category. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Differentiation along the tail

	1^{st} to 10^{th} percentile (1)	10^{th} to 25^{th} percentile (2)	25^{th} to 50^{th} percentile (3)	50^{th} to 75^{th} percentile (4)	75^{th} to 100^{th} percentile (5)	Sum of affiliations (6)
$D_i * post_t$	-0.200***	-0.102***	-0.104***	-0.068**	0.081**	-0.376***
	(0.0420)	(0.0380)	(0.0400)	(0.0356)	(0.0315)	(0.0959)
First stage	0.0292***	0.0306***	0.0306***	0.0306***	0.0306***	0.031***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
F-test of excluded instruments	152.86	166.70	166.70	166.70	166.70	169.44
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X
YouTuber Time Trend	X	X	X	X	X	X
YouTubers	10,599	10,591	10,591	10,591	10,591	10,589
Videos	1,064,248	1,033,666	1,033,666	1,033,666	1,033,666	1,028,446

Notes: Robust standard errors in parentheses. Each column displays the results of a 2SLS estimation. In column 1, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 1^{st} to 10^{th} percentile of the distribution of most-viewed keywords. In column 2, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 10^{th} to 25^{th} percentile of the distribution of most-viewed keywords. In column 3, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 25^{th} to 50^{th} percentile of the distribution of most-viewed keywords. In column 4, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 50^{th} to 75^{th} percentile of the distribution of most-viewed keywords. In column 5, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 75^{th} to 100^{th} percentile of the distribution of most-viewed keywords. In column 6, the dependent variable is the sum of a video's percentile indicators. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Mechanism: Competitive content

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	-0.003	-0.003	-0.005	-0.212***	-0.210***	-0.179***			
	(0.0075)	(0.0074)	(0.0073)	(0.0481)	(0.0477)	(0.0456)			
$close_i * post_t$							-0.006***	-0.006***	-0.005***
							(0.0013)	(0.0013)	(0.0012)
First stage				0.0286***	0.0286***	0.0290***			
				(0.0023)	(0.0023)	(0.0024)			
F-test of excluded instruments				144.13	143.85	151.32			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Competitive_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a competitive keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Competitive content – Evidence from the video level

	Competitive (1)	Competitive (2)	Competitive (3)	Non-comp. (4)	Non-comp. (5)	Non-comp. (6)
$close_i * post_t$	0.0011	0.0011	0.0010	0.003**	0.003**	0.003**
	(0.0010)	(0.0010)	(0.0010)	(0.0014)	(0.0014)	(0.0013)
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	10,332	10,332	10,332	9,550	9,550	9,550
Videos	693,449	693,449	693,449	373,444	373,444	373,444

Notes: Robust standard errors in parentheses. The dependent variable is $I(\geq 10)_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is ten minutes or longer, and 0 otherwise. All estimates are obtained by OLS and based on using the advertising YouTubers only. In addition, the estimates in Columns 1 to 3 are based on videos classified as competitive. The estimates in Columns 4 to 6 are based on videos classified as non-competitive. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Video duration and competitive content

	OLS (1)	OLS (2)	OLS (3)
video duration	-0.0000298	-0.0000666	-0.0000695
	(0.000116)	(0.000114)	(0.0001139)
Time FE	X	X	X
YouTuber FE	X	X	X
Controls		X	X
YouTuber Time Trend			X
YouTubers	10,113	10,113	10,113
Videos	566,079	566,079	566,079
R^2	0.404	0.412	0.3878

Notes: Robust standard errors in parentheses. The dependent variable is $Competitive_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is classified as competitive, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, the estimates are based on a regression that excludes all months $t \geq 34$. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Commentator analysis

	OLS (1)	2SLS (2)	Red. Form (3)
D_i	-0.0142***	-0.0475**	
	(0.0038)	(0.0232)	
$close_i$			-0.0014**
			(0.0006)
First stage		0.0289***	
		(0.0023)	
F-test of excluded instruments		159.78	
YouTubers	5,907	5,907	5,907

Notes: Robust standard errors in parentheses. The dependent variable is $\Delta fluctuation_i$, which is the difference in the commentator fluctuation before and after Oct 2015 for YouTuber i. The estimates are based on the advertising YouTubers only. Only YouTubers who received more than 25 comments before and after Oct 2015 are included in the analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Commentator analysis – Non-advertising YouTubers

	OLS (1)	2SLS (2)	Red. Form (3)
D_i	-0.022***	-0.039	
	(0.008)	(0.055)	
$close_i$			-0.0009
			(0.0013)
First stage		0.024***	
		(0.004)	
F-test of excluded instruments		34.33	
YouTubers	1,462	1,462	1,462

Notes: Robust standard errors in parentheses. The dependent variable is $\Delta fluctuation_i$, which is the difference in the commentator fluctuation before and after Oct 2015 for YouTuber i. The estimates are based on the non-advertising YouTubers only. Only YouTubers who received more than 25 comments before and after Oct 2015 are included in the analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Quality - Likes / (Likes + Dislikes)

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	-0.0001	0.0001	0.0001	-0.041***	-0.040***	-0.039***			
	(0.0015)	(0.0015)	(0.0108)	(0.0015)	(0.0108)	(0.0107)			
$close_i * post_t$							-0.001***	-0.001***	-0.001***
							(0.0003)	(0.0003)	(0.0003)
First stage				0.028***	0.028***	0.028***			
				(0.0025)	(0.0024)	(0.0024)			
F-test of excluded instruments				130.86	130.60	137.57			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,594	10,594	10,594	10,594	10,594	10,594	10,594	10,594	10,594
Videos	990,476	990,476	990,476	990,476	990,476	990,476	990,476	990,476	990,476

Notes: Robust standard errors in parentheses. The dependent variable is $\frac{Likes}{Likes+Dislikes}_{vit}$, i.e., the share of positive ratings for video v of YouTuber i in month t. The estimates are based on using the advertising YouTubers only. Videos that received no likes nor dislikes are excluded from the analysis. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Quality - log(Views)

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	0.207***	0.210***	0.216***	0.436***	0.469***	0.297*			
	(0.0201)	(0.0261)	(0.0260)	(0.1586)	(0.1592)	(0.1518)			
$close_i * post_t$							0.0125***	0.0134***	0.0086**
							(0.0044)	(0.0044)	(0.0044)
First stage				0.0285***	0.0286***	0.0290***			
				(0.0023)	(0.0023)	(0.0024)			
F-test of excluded instruments				144.13	143.85	151.44			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $log(Views)_{vit}$, which is the logarithm of the views video v of YouTuber i uploaded in month t has received. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Alternative observation period

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. F. (7)	Red. F. (8)	Red. F. (9)
$D_i * post_t$	0.008 (0.008)	0.005 (0.008)	0.005 (0.008)	-0.131*** (0.048)	-0.130*** (0.047)	-0.124***			
$close_i * post_t$							-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)
First stage				0.029*** (0.002)	0.029*** (0.002)	0.029*** (0.002)			
F-test of excluded instruments				147.47	147.48	153.11			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,513	10,513	10,513	10,513	10,513	10,513	10,513	10,513	10,513
Videos	745,219	745,219	745,219	745,219	745,219	745,219	745,219	745,219	745,219

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only the periods $t \in [13, 43]$ are included into the analysis. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 19: Alternative selection of YouTubers – median video duration < 7.5

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. F. (7)	Red. F. (8)	Red. F. (9)
$D_i * post_t$	0.007 (0.008)	0.005 (0.008)	0.005 (0.008)	-0.246*** (0.050)	-0.240*** (0.049)	-0.234*** (0.048)			
$close_i * post_t$							-0.010*** (0.002)	-0.010*** (0.002)	-0.009*** (0.002)
First stage				0.040*** (0.003)	0.040*** (0.003)	0.040*** (0.003)			
F-test of excluded instruments				161.21	160.26	166.53			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	9,519	9,519	9,519	9,519	9,519	9,519	9,519	9,519	9,519
Videos	923,189	923,189	923,189	923,189	923,189	923,189	923,189	923,189	923,189

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers whose median video duration before Oct 2015 is smaller than 7.5 are included into the analysis. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 20: Alternative selection of YouTubers – 90^{th} percentile < 10 min

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. F. (7)	Red. F. (8)	Red. F. (9)
$D_i * post_t$	0.009	0.005	0.004	-0.213***	-0.234***	-0.234***			
	(0.011)	(0.010)	(0.011)	(0.057)	(0.056)	(0.057)			
$close_i * post_t$							-0.011***	-0.012***	-0.012***
							(0.003)	(0.002)	(0.003)
First stage				0.054***	0.053***	0.053***			
				(0.005)	(0.005)	(0.005)			
F-test of excluded instruments				135.13	134.03	134.40			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	6,891	6,891	6,891	6,891	6,891	6,891	6,891	6,891	6,891
Videos	610,496	610,496	610,496	610,496	610,496	610,496	610,496	610,496	610,496

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers whose 90^{th} percentile of the distribution of video durations before Oct 2015 is smaller than 10 are included into the analysis. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 21: Alternative classifications of the treatment group – cutoffs

	1% (1)	1% (2)	1% (3)	10% (4)	10% (5)	10% (6)
$D_i * post_t$	-0.245***	-0.234***	-0.232***	-0.362***	-0.347***	-0.329***
	(0.060)	(0.060)	(0.060)	(0.090)	(0.087)	(0.084)
First stage	0.024 ***	0.024***	0.024***	0.016***	0.016***	0.017***
	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
F-test of excluded instruments	73.31	73.33	71.54	92.66	92.89	100.79
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. All estimates are 2SLS estimates. In columns 1 to 3, YouTubers who have increased their share of videos between ten and fourteen minutes by at least 1 percentage point after Oct 2015 are classified as treated. Analogously, in columns 4 to 6, YouTubers who have increased their share of videos between ten and fourteen minutes by at least 10 percentage point after Oct 2015 are classified as treated. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 22: Alternative classifications of the treatment group – all videos $\geq$ 10 minutes

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. F. (7)	Red. F. (8)	Red. F. (9)
$D_i * post_t$	0.005	-0.0001	-0.0008	-0.151***	-0.145***	-0.141***			
	(0.007)	(0.007)	(0.007)	(0.036)	(0.035)	(0.035)			
$close_i * post_t$							-0.005***	-0.006***	-0.006***
							(0.001)	(0.001)	(0.001)
First stage				0.039***	0.039***	0.039***			
				(0.003)	(0.003)	(0.003)			
F-test of excluded instruments				207.91	208.62	210.86			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, YouTubers who have increased their share of videos that are ten minutes or longer by at least five percentage points are classified as treated. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 23: Alternative definitions of the instrument

	75^{th} perc. (1)	75^{th} perc. (2)	75^{th} perc. (3)	90^{th} perc. (4)	90^{th} perc. (5)	90^{th} perc. (6)
$D_i * post_t$	-0.131*	-0.148**	-0.140*	-0.376**	-0.425***	-0.412***
	(0.074)	(0.074)	(0.073)	(0.147)	(0.161)	(0.157)
First stage	0.007***	0.007***	0.007***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.0005)	(0.0006)	(0.0006)
F-test of excluded instruments	24.71	24.58	24.83	14.68	14.34	14.71
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. All estimates are 2SLS estimates. In columns 1 to 3, the instrument $close_i$ is defined as the 75^{th} percentile in the distribution of video durations of YouTuber i before Oct 2015. In columns 4 to 6, the instrument $close_i$ is defined as the 90^{th} percentile in the distribution of video durations of YouTuber i before Oct 2015. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 24: Alternative cutoffs for the commentator analysis

	> 100	> 100	> 100	> 50	> 50	> 50	> 10	> 10	> 10	> 0	> 0	> 0
	OLS	2SLS	Red. Form	OLS	2SLS	Red. Form	OLS	2SLS	Red. Form	OLS	2SLS	Red. Form
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
D_i	-0.009**	-0.066***		-0.01**	-0.050**		-0.017***	-0.035		-0.035***	0.038	
	(0.004)	(0.024)		(0.004)	(0.022)		(0.004)	(0.025)		(0.005)	(0.031)	
$close_i$			-0.002***			-0.002**			-0.0010			0.001
			(0.0007)			(0.0007)			(0.0459)			(0.0008)
First stage		0.031***			0.031***			0.028***			0.028***	
		(0.0028)			(0.0025)			(0.0021)			(0.0019)	
F-test of excluded instruments		120.47			150.96			176.90			221.06	
YouTubers	3,989	3,989	3,989	4,924	4,924	4,924	7,098	7,098	7,098	8,916	8,916	8,916

Notes: Robust standard errors in parentheses. The dependent variable is $\Delta fluctuation_i$, which is the difference in the commentator fluctuation before and after Oct 2015 for YouTuber i. The estimates are based on the advertising YouTubers only. In columns 1 to 3, only YouTubers who received more than 100 comments before and after Oct 2015 are included in the analysis. In columns 4 to 6, only YouTubers who received more than 50 comments before and after Oct 2015 are included in the analysis. In columns 7 to 9, only YouTubers who received more than 10 comments before or after Oct 2015 are included in the analysis. In columns 10 to 12, all YouTubers who received at least one comment before and after Oct 2015 are included in the analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 25: Video duration and likes/(likes+dislikes) and log(Views)

	$\frac{Likes}{Likes+Disl}$	$\frac{Likes}{Likes+Disl}$	$\frac{Likes}{Likes+Disl}$	$log(Views)$	$log(Views)$	$log(Views)$
video duration	0.000109*** (0.000035)	0.000105*** (0.000035)	0.000109*** (0.000035)	0.0146*** (0.0013)	0.0146*** (0.0013)	0.0146*** (0.0013)
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	10,068	10,068	10,068	10,111	10,111	10,111
Videos	518,166	518,166	518,166	565,963	565,963	565,963
R^2	0.297	0.297	0.297	0.697	0.697	0.704

Notes: Robust standard errors in parentheses. In columns 1 to 3, the dependent variable is $Likes/(Likes + Dislikes)_{vit}$. In columns 4 to 6, the dependent variable is $log(Views)_{vit}$. The estimates are based on using the advertising YouTubers only. Moreover, the estimates are based on an OLS regression that excludes all months $t \geq 34$. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 26: Correlation measures popular content

	1%	0.5%	2%	Fixed	Likes
1%	1.0000				
0.5%	0.8237	1.0000			
2%	0.8207	0.6761	1.0000		
Fixed	0.8495	0.8479	0.7528	1.0000	
Likes	0.8298	0.7688	0.7802	0.7598	1.0000

Notes: Correlation matrix for the different measures of popular content.

Table 27: Alternative definitions of popular content

	0.5% (1)	0.5% (2)	0.5% (3)	2% (4)	2% (5)	2% (6)	Fixed (7)	Fixed (8)	Fixed (9)	Likes (10)	Likes (11)	Likes (12)
$D_i * post_t$	-0.029	-0.0198	-0.0198	-0.361 ***	-0.355***	-0.343***	-0.142***	-0.143***	-0.133***	-0.252***	-0.231***	-0.232***
	(0.0464)	(0.0451)	(0.0442)	(0.0555)	(0.0545)	(0.0533)	(0.0477)	(0.0469)	(0.0459)	(0.0519)	(0.0503)	(0.0496)
First stage	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
F-test of excluded instruments	144.13	143.85	151.32	144.13	143.85	151.32	144.13	143.85	151.32	144.13	143.85	151.32
Time FE	X	X	X	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X		X	X
YouTuber Time Trend			X			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. In columns 1 to 3, the dependent variable $Popular_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper half percent of the distribution of most-viewed keywords. In columns 4 to 6, the dependent variable $Popular_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper two percent of the distribution of most-viewed keywords. In columns 7 to 9, the dependent variable $Popular_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from a fixed number of the distribution of most-viewed keywords. In columns 10 to 12, the dependent variable $Popular_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper one percent of the distribution of most-liked keywords. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 28: Correlation measures competitive content

	1%	0.5%	2%	Fixed
1%	1.0000			
0.5%	0.8424	1.0000		
2%	0.8415	0.7093	1.0000	
Fixed	0.8602	0.8552	0.7819	1.0000

Notes: Correlation matrix for the different measures of competitive content.

Table 29: Alternative definitions of competitive content

	0.5% (1)	0.5% (2)	0.5% (3)	2% (4)	2% (5)	2% (6)	Fixed (7)	Fixed (8)	Fixed (9)
$D_i * post_t$	-0.232***	-0.231***	-0.208***	-0.178***	-0.176***	-0.143***	-0.176***	-0.183***	-0.154***
	(0.0519)	(0.0514)	(0.0494)	(0.0436)	(0.0434)	(0.0416)	(0.0483)	(0.0483)	(0.0465)
First stage	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
F-test of excluded instruments	144.13	143.85	151.32	144.13	143.85	151.32	144.13	143.85	151.32
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. In columns 1 to 3, the dependent variable $Competitive_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper half percent of the distribution of most-used keywords. In columns 4 to 6, the dependent variable $Competitive_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper two percent of the distribution of most-used keywords. In columns 7 to 9, the dependent variable $Competitive_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from a fixed number of the distribution of most-used keywords. All estimates are based on 2SLS regressions. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 30: YouTube platform events

	Date		Summary of the event	Moneti-zation	Content choice
1	2013	Jan	The channel view count only includes views from publicly available videos from now.		
2	2013	Feb	It is now technically feasible to update several video updates at the same time.		
3	2013	Mar	YouTube changes the interaction with AdSense: a YouTuber's financial overview is now available at YouTube Analytics	X	
4	2013	Mar	The new channel design "YouTube One" is available for all YouTubers.		
5	2013	Apr	Users see more videos in their homepage feed.		
6	2013	May	YouTubers receive an e-mail once a video upload has finished.		
7	2013	May	The new channel design "YouTube One" is mandatory for all YouTubers.		
8	2013	May	Selected YouTubers from the US may raise a subscription fee of 0.99$ per month.	X	X
9	2013	June	Mobile users (Android and iOS) may follow links embedded into videos from now.		
10	2013	July	YouTubers may now connect multiple channels via a Google+ page.		
11	2013	Aug	Improved mobile features for users.		
12	2013	Sept	Launch of the YouTube Audio Library (150 royalty-free tracks).		
13	2013	Sept	Improved tools for moderating comments.		
14	2013	Sept	New tools to identify and interact with one's top viewers.		
15	2013	Sept	YouTubers may now feature playlists from other channels.		
16	2013	Oct	Selected YouTubers from outside the US may also raise a subscription fee of 0.99$ per month.	X	X
17	2013	Nov	A YouTuber may let the platform turn her video into an ad that is then shown to viewers from different channels.		X
18	2013	Dec	Live streams are now technically feasible.		X
19	2014	Feb	YouTube validates a videos view count repeatedly from now on.		
20	2014	Feb	Users can create their own playlists.		
21	2014	Apr	Enhanced playlist tools in YouTube Analytics are launched.		
22	2014	June	New messaging and commenting features for YouTubers.		
23	2014	June	YouTube removes blocked users from a channel's subscriber count.		
24	2014	Nov	New YouTube homepage for music videos.		
25	2015	Mar	360 degree videos are now technically feasible.		X
26	2015	May	60fps for live streams is now technically feasible.		
27	2015	June	New data tool Music Insights is available: shows the cities where an artist is most popular, top tracks by artist, and views from both artists official music videos and fan uploads claimed using Content ID.		
28	2015	July	A new design for YouTube mobile app is launched.		
29	2015	Oct	YouTube Red is launched in the US.	X	X
30	2015	Nov	New language and translation tools are available.		
31	2015	Nov	New virtual reality tools are available.		X
32	2016	Jan	Users can donate to the YouTuber after watching a video.	X	X
33	2016	Feb	A new blurring tool (to blur faces etc.) is available.		
34	2016	Apr	YouTube withholds any ad revenue generated during content ID disputes from now.	X	
35	2016	June	Mobile live streams are now technically feasible.		X
36	2016	Sept	YouTube Analytics becomes easier to understand for YouTubers.		
37	2016	Sept	New tools for YouTubers to engage with their community.		
38	2016	Oct	An optional feature for paid promotion disclosure is available.	X	X
39	2016	Oct	Special video end screens are available.		X
40	2016	Nov	New comment features are available for users.		
41	2016	Dec	Launch of a new URL system that is independent from Google+.		
42	2017	Jan	User messages in a chat stream may be highlighted.		

Notes: Summary of YouTube platform events.

Table 31: Learning

	Popular (1)	Popular (2)	Popular (3)	Competitive (4)	Competitive (5)	Competitive (6)
$D_i * post_t$	-0.192***	-0.020***	-0.196***	-0.179***	-0.188***	-0.185***
	(0.0480)	(0.0485)	(0.0482)	(0.0014)	(0.0460)	(0.0458)
t_i	-0.00008	0.0004***	0.0007***	-0.0003***	0.0002*	0.0005***
	(0.00005)	(0.0001)	(0.0001)	(0.00004)	(0.0001)	(0.0001)
t_i^2		-1.13e-06***	-3.62e-06***		-1.24e-06***	-3.36e-06***
		(2.54e-07)	(8.93e-07)		(2.52e-07)	(9.02e-07)
t_i^3			4.50e-09***			3.82e-09**
			(1.69e-09)			(1.71e-09)
First stage	0.0290***	0.0288***	0.0289***	0.0290***	0.0288***	0.0289***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
F-test of excluded instruments	151.32	148.55	150.05	151.32	148.55	150.05
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. In columns 1 to 3, the dependent variable is $Popular_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. In columns 4 to 6, the dependent variable is $Competitive_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a competitive keyword, and 0 otherwise. All estimates are obtained by 2SLS and based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$