CESifo
AREA
CONFERENCES 2021

Economics of Digitization
Munich, 19 – 20 November 2021

Advertising and Content Differentiation: Evidence from YouTube

Anna Kerkhof

CESifo

Advertising and Content Differentiation: Evidence from YouTube[*]

Anna Kerkhof[†]

April 14, 2021

Abstract

Many media outlets depend on advertising revenue to finance their operations, but the causal effect of advertising on media outlets' content choice is largely unexplored. This paper exploits two institutional features from YouTube to show that an exogenous increase in advertising quantity induces YouTubers to differentiate their video content from their competitors. A plausible mechanism is that YouTubers wish to avoid competition. Viewers typically perceive advertising as a nuisance and therefore as an implicit price they have to pay, thus, they could switch to a competitor if a YouTuber increased her advertising quantity. This is less likely, however, if the YouTuber differentiates her content from the mainstream, gains market power in a niche, and thereby softens competition in the ad "price."

JEL Codes: D22, L15, L82, L86
Keywords: advertising, content differentiation, economics of digitization, horizontal product differentiation, long tail, media diversity, user-generated content, YouTube

1. Introduction

Media consumption is an integral part of our everyday lives. American adults spend several hours per day reading, watching, and listening to media content[1]; similar numbers hold for citizens from the EU.[2] The value of media consumption crucially depends on the match between consumers' preferences and the media content that is being provided. Preferences over media content, however, differ substantially between different groups of consumers: men and women, for instance, prefer very different types of media content (Anderson and Waldfogel, 2015). Thus, the more horizontally differentiated the content in media markets, the more likely it is that all consumers' preferences are served and that their value of media consumption is maximized (Waldfogel, 2007). Duplication of media content, in contrast, may lead to foregone consumer surplus, business stealing, and the inefficient duplication of fixed costs.[3]

What drives content differentiation in media markets? The extensive literature on horizontal product differentiation has singled out two general effects (Tirole, 1988; Anderson et al., 1992). First, a direct effect induces firms to move closer to their competitors to increase their market share, leading to minimum differentiation (Hotelling, 1929). Second, a strategic effect prompts firms to move away from their competitors to soften price competition, leading to maximum differentiation (d'Aspremont et al., 1979; Economides, 1984, 1986) – when prices are fixed, the direct effect prevails. Media outlets, however, typically charge their consumers a low or even no price at all and generate revenue through advertising instead, provoking a persistent debate about the effect of advertising on content differentiation in media markets. Pioneering models on media outlets' content choice by Steiner (1952), Beebe (1977), and Gabszewicz et al. (2001) argue that, in the absence of price competition, advertising leads to minimum content differentiation, as media outlets strive to sell a maximum number of eyeballs to advertisers. More recent models, on the other hand, 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 second-most visited website in the world[4] – to resolve the open question. I demonstrate that an exogenous increase in the feasible advertising quantity leads to a considerable decrease in the YouTubers' probability to duplicate mainstream content, i.e., the type of content that attracts the largest number of views. The result is driven by an intuitive

---

[1]See https://www.nielsen.com/us/en/insights/article/2018/time-flies-us-adults-now-spend-nearly-half-a-day-interacting-with-media/ (Aug 2020).

[2]See https://www.emarketer.com (Aug 2020).

[3]See also Baker and George (2019), who analyze product positioning and welfare in the market for local TV news.

[4]See www.alexa.com/topsites (Aug 2020).

mechanism: Mainstream content is provided by many competing YouTubers; thus, viewers could easily switch to a competitor if a YouTuber increased her advertising quantity. Switching is less likely, however, if the YouTuber differentiates her content from the mainstream, gains market power in a niche, and thereby softens competition in the ad "price." The empirical findings therefore support models that acknowledge a conceptual equivalence of advertising and subscription prices as determinants of content differentiation in media markets. Moreover, the results are in line with more general predictions about the interplay of price competition and horizontal product differentiation: When prices are fixed (or advertising quantities are limited), firms locate close to their competitors. If price competition is possible (if the feasible advertising quantity goes up), the strategic effect becomes stronger, whereby product differentiation increases.

I exploit two institutional features of YouTube to identify causal effects. First, I use 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 unknown to the majority of YouTubers until Nov 2015, when YouTube launched a new ad break tool that made its existence prominent to the community. I focus on a subsample of YouTubers who were likely to be unaware of the ten minutes trick before Nov 2015; in other words, I consider a subsample of YouTubers whose advertising quantity was limited. Then, I compare the change in the probability to duplicate mainstream content of YouTubers whose feasible advertising quantity increased after Nov 2015 (YouTubers who increased their share of videos that are ten minutes or longer) with the change in content of YouTubers whose feasible advertising quantity remained constant in a difference-in-differences framework.

The increase in the feasible advertising quantity is endogenous, though – after all, the YouTubers have perfect control over their videos' duration, and particularly money-loving YouTubers may be especially eager to produce longer videos to show more ads. To take this into account, I use the YouTubers' *median video duration before Nov 2015* – in a sense, their "closeness" to the ten minutes threshold – to instrument for the increase in the feasible advertising quantity. I argue that the instrument is relevant and exogenous. On the one hand, a YouTuber's median video duration before Nov 2015 is correlated to the potential increase in the feasible number of ad breaks per video, as extending the videos' duration to at least ten minutes is easier for YouTubers with videos close to the threshold. On the other hand, the YouTubers in my sample did not choose their videos' duration before Nov 2015 bearing the ten minutes trick in mind, since they were unaware of the feature. As a result, the median video duration before Nov 2015 is exogenous to omitted variables that might drive an increase in the feasible advertising quantity (e.g., commercial interests) and has furthermore no effect on the YouTubers' probability to duplicate mainstream content except through the potential increase in advertising quantity. A broad range of validity checks supports the identification strategy.

3

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 mainstream 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 easier (e.g., a music clip is typically between three and five minutes long and cannot be easily extended).

In line with the literature on horizontal product differentiation and, in particular, with recent models on content choice in media markets (see Anderson and Jullien, 2016, for a survey), I also demonstrate that softening competition in the ad "price" is a plausible economic mechanism behind my main results. As a first step, I show that mainstream content – i.e., content in high demand – is also *provided* by many YouTubers. Thus, YouTubers who duplicate mainstream content run the risk of losing their audience to a competitor if they increase their advertising quantity. This is less likely, however, if the YouTuber differentiates her content from the mainstream, gains market power in a niche, and thereby softens competition in the ad "price." Thus, as a second step, I show that an increase in the feasible advertising quantity leads to a twenty percentage point reduction in the YouTubers' probability to upload content that is covered by many other YouTubers, too. I further support the argument by studying the YouTubers' viewers themselves. If YouTubers differentiate their content from the mainstream to prevent their viewers from switching to a competitor, viewer fluctuation should go down, which is indeed the case. I find no evidence for other economic mechanisms behind my results.

My paper makes five contributions to the literature. First, I advance the knowledge 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 – public and academic – opinion that media outlets duplicate mainstream content when incentivized by ad revenue (e.g., Herman and McChesney, 1997; Hamilton, 2004; McChesney, 2004). This is a major insight, especially because the media's options to generate ad revenue are often subject to external regulation.[5] My paper shows that upper bounds on advertising quantity may reduce content differentiation in media markets. It is, however, widely agreed that a diverse media landscape helps to sustain democracy, freedom, and the public discourse (Downs, 1957; Coase, 1974; Gentzkow et al., 2015; Puglisi and Snyder, 2015); thus, policy makers undertake great efforts to achieve and maintain media markets with a high degree

---

[5]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).

of content differentiation.[6] My results suggest that upper bounds on advertising may counteract such endeavors, and regulation authorities should take this into account.

Second, I provide empirical evidence of the conceptual equivalence of advertising and subscription prices as determinants of content differentiation in media markets. 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, my paper is among the first to show that increases in advertising quantity lead to an analogous effect. Closely related is the work by 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 mainstream content. Our results complement each other: 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. Relatedly, my paper amplifies the broader empirical literature on the interplay of price competition and horizontal product differentiation, including, e.g., Numan and Willekens (2012), who show that auditors compete on fees and specialize in certain industries to soften price competition, (Netz and Taylor, 2002), who show that gasoline stations in the US spatially differentiate as a result of price competition, and (Davcik and Sharma, 2015), who show that product differentiation in terms of branding is associated with higher prices.

Third, 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."[7] 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 a large number of media outlets could in principle co-exist, they might duplicate the most mainstream 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.

Forth, the paper is linked to analyses on further determinants of content differentiation in media markets. Petrova (2011), for instance, shows that in the nineteenth century, US newspapers with higher advertising revenues were more likely to be independent of politi-

---

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

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

cal parties. Several papers consider the effect of mergers and ownership concentration on product positioning and product variety in media markets. Berry and Waldfogel (2001) and Sweeting (2010) show that ownership concentration in the US radio market increased the number of formats relative to the number of stations, most likely to avoid business stealing effects stations with common owners. Similarly, George (2007) finds that ownership concentration in US newspaper markets increases differentiation and the variety of topics.

Finally, 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 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 competition as a plausible economic mechanism driving 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).[8] As of August 2020, more than a billion hours of video content from YouTube are watched every day.[9]

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.[10]

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.[11] According to anecdotal evidence – official statistics do not exist – YouTubers earn about three to five USD per 1,000 views per ad per video.[12]

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.[13] 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.[14] 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.[15] 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.[16]

---

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

[9]See `www.youtube.com/yt/about/press/` (Aug 2020).

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

[11]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.

[12]See `influencermarketinghub.com/how-much-do-youtubers-make/` (Dec 2018).

[13]See `support.google.com/youtube/answer/6162278?hl=en` (Dec 2018).

[14]See `nymag.com/intelligencer/2017/12/can-youtube-survive-the-adpocalypse.html` (Dec 2018).

[15]See `turbofuture.com/internet/YouTube-Screwed-Small-YouTube-Channels-With-Their-New-Memorization-Policy` (Dec 2018).

[16]`support.google.com/youtube/answer/72857?hl=en\&ref_topic=6029709` (Dec 2018).

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.[17] 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 Nov 2015, when YouTube launched a new ad break tool for YouTubers.[18] 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 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 mainstream content. Given that the YouTubers' real life monetization settings are endogenous, however, the identification of a causal link from advertising quantity to individual 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* Nov 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

---

[17]`support.google.com/youtube/answer/6175006?hl=en` (Oct 2018).
[18]See `www.youtube.com/watch?v=z58Ed6q6xQg` (Oct 2018).

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 Nov 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 costs of moving beyond the threshold *after* it became prominent – and thereby also the probability to actually do so – is exogenous to unobserved characteristics such as, for instance, commercial incentives that may also drive a YouTuber's decision to increase her feasible advertising quantity.[19]

I exploit the variation in the YouTubers' costs to move beyond the threshold as follows. First, I consider only YouTubers like A and B, i.e., YouTubers below ten minutes threshold before Nov 2015, whose feasible advertising quantity was therefore restricted. Then, I compare the change in the probability to upload mainstream content before and after Nov 2015 of YouTubers who could increase the feasible number of ad breaks per video by uploading videos that are ten minutes or longer to YouTubers whose feasible advertising quantity remained constant in a difference-in-differences framework. Finally, I account for endogeneity in the increase of feasible advertising quantity by using a YouTuber's "closeness" to the ten minutes threshold before Nov 2015 as an instrument. Thus, my empirical strategy boils down to exploiting exogenous variation between YouTubers who were close to the threshold before Nov 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 `channelcrawler.com` to compile a list of all active German YouTube channels as of Oct 2017. Based on this list, I collect data on the

---

[19]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 instrument heterogeneity.

YouTuber level, including a full history of video uploads by each YouTuber, from the YouTube 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 a snapshot 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."[20] Hence, although one could principally detect the number of permitted ad breaks in a video's HTML code, a webscraper could not crawl each video in the dataset within a reasonable amount of time (it would take years). To retrieve data on the YouTubers' monetization settings nonetheless, I make the following compromise: I let a webscraper crawl twenty randomly drawn videos per YouTuber. 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.[21] This 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. As a consequence, the subsequent analysis does not study the relationship between the *actual* number of ad breaks per video and the probability to duplicate mainstream content. Rather – building on the deterministic link between video duration and the *feasible* number of ad breaks per video (see Section 2) – my analysis unveils how the *option* to increase the number of ad breaks per video affects the probability to duplicate mainstream content. The relevance of this question is further discussed in Section 5.2.1; in addition, I discuss the consequences of a potential measurement error in Appendix D.1.

4.2. Definition of mainstream content

Following the procedure in Sun and Zhu (2013), I use the number of video views and the videos' keywords to construct a measure for mainstream 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.[22] 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 "mainstream."[23] Finally, I assign a dummy variable that is equal to one to all videos equipped with a mainstream keyword. Note that it is important to consider each month and each video category separately. First,

---

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

[21]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 to four months.

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

[23]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."

what is mainstream 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 mainstream 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 mainstream 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 Nov 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 demonetization 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 C.1 and Appendix D.2.

Second, I determine which YouTubers to include. Following the outline from Section 3, I restrict the analysis to YouTubers below the ten minutes threshold before Nov 2015 (YouTubers A and B in the example), where I use a YouTuber's median video duration before Nov 2015 to define her "position" on the x-axis in Figure 3. Thus, I include only YouTubers whose median video duration before Nov 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 Nov 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 C.2 shows several robustness checks based on different selections of YouTubers.

4.4. Illustrative evidence

Based on the final dataset from Section 4.3, this section provides illustrative evidence of the three main arguments from the stylized example in Section 3. That is, I confirm that the YouTubers whom I selected were unaware of the ten minutes trick before Nov 2015, 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 Nov 2015 were more likely to exploit it. I also provide video level evidence of an increase in the *actual* (not feasible) number of ad breaks per video.

Figure 5 shows the development of the share of videos between ten and fourteen minutes for advertising and non-advertising YouTubers.[24] If the launch of the new ad break tool made the ten minutes trick more apparent, this share should increase for advertising, but not for non-advertising YouTubers after Nov 2015, which is indeed the case.

As a further comparison of advertising and non-advertising YouTubers may be problematic – after all, they act based on entirely different motives – the remainder of the paper focuses on advertising YouTubers. Figure 6 illustrates that (advertising) YouTubers who were closer to the ten minutes threshold before Nov 2015 (YouTubers like B in the stylized example) are indeed more likely to increase their share of videos between ten and fourteen minutes after Nov 2015 than YouTubers who were further away from it (YouTubers like A in the example). Since "closeness" – in terms of a YouTuber's median video duration before Nov 2015 – is a continuous measure, I cannot compare the trends of distinct groups, though. Instead, Figure 6 shows that the increase in the share of videos between ten and fourteen minutes after Nov 2015 is steeper for YouTubers around the 75th percentile of median video duration before Nov 2015 than for YouTubers around the 25th percentile.

Next, I examine the distribution of video durations before and after Nov 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 Nov 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* Nov 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 Nov 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 Nov 2015, the bunching should be more pronounced for YouTubers with

---

[24]YouTubers who exploit the ten minutes trick to benefit from the increase in the feasible number of ad breaks per video are likely to produce videos that are *just* longer than ten minutes. Considering the share of videos between ten and fourteen minutes is thus more informative than considering the share of all videos that are ten minutes or longer.

a higher median video duration before Nov 2015; Figures 8 and 9 confirm that this is the case. Third, the distributions of video durations before Nov 2015 in Figures 7 and 9 document that the YouTubers in my sample were indeed ignorant of the ten minutes trick before the launch of the new ad break tool, because – 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 (Mc-Crary, 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. Prior to Nov 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 Nov 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).[25] 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 Nov 2015 to 1.04 after Nov 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 Nov 2015, this number has risen to 52 after Nov 2015. Thus, in the random subsample, the actual number of ad breaks has increased on the intensive as well as on the extensive margin.

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 her feasible advertising quantity after Nov 2015. Thus, the treatment corresponds to gaining the *option* to show more ads than before, rather than actually increasing in the number of ad breaks. In other words, I study the effect of being liberated from an upper bound on advertising quantity on the decision to duplicate mainstream content or not; see Section 5.2.1 for further discussion. To this end, I compute each YouTuber's share of videos between ten and fourteen minutes[26] before and after Nov 2015; if this share has increased by at least

---

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

[26]See Section 4 for a discussion on why considering the interval from ten to fourteen minutes is appropriate.

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 C.3 for robustness checks that use other classifications of treatment and control group.

The baseline difference-in-differences regression is given by

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

where D_i indicates the treatment group, $post_t$ indicates all months after Nov 2015, X_{vit} controls for video categories, ϕ_i and ϕ_t are YouTuber and monthly fixed effects, respectively, and t_{it} is a YouTuber specific linear time trend. The dependent variable $Mainstream_{vit}$ is a dummy variable equal to one if video v of YouTuber i in month t is given a mainstream 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 duplicate mainstream 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 an increase in the feasible advertising quantity on the YouTubers' probability to duplicate mainstream 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 duplicate mainstream 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 $Mainstream_{vit}$ 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 duplicate mainstream content, the OLS estimate of β would, again, be upward biased. Finally, reverse causality may generate a spurious relationship between $Mainstream_{vit}$ and D_i. If, for instance, YouTubers who produce more mainstream 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 the YouTubers' treatment status, I use YouTubers' *median video duration before Nov 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 Nov 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) by Two Stage Least Squares (2SLS).

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

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

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 YouTubers' probability to duplicate mainstream content before and after Nov 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 mainstream content, whereby it is comparable to an intention-to-treatment effect. In contrast to that, β corresponds to the average effect of gaining the concrete option to have more ad breaks on the difference in the probability to duplicate mainstream content before and after Nov 2015.

Note that the IV approach isolates exogenous variation in the YouTubers' *feasible*, but not in their *actual* number of ad breaks per video. While variation in $close_i$ is likely to affect whether a YouTuber gains the *option* to permit for additional ad breaks (as captured in D_i), the instrument is unlikely to be correlated with the *actual number* of additional ad breaks a YouTuber chooses. In other words, $close_i$ is unlikely to affect whether a YouTuber permits for one, two, or even more additional ad breaks per video. Thus, I can identify the causal effect of an increase in the feasible number of ad breaks per video, but must stay agnostic about the causal marginal effect of each additional ad.[27] I do not perceive this as a severe shortcoming of the analysis, though. First, considering a YouTuber's content choice on the left- and a video's actual number of ad breaks on the right-hand-side of the equation would raise simultaneity issues. Studying an exogenous increase in the option to have additional ad breaks, in contrast, does not involve such concerns. Second, YouTubers can modify their monetization settings at any point in time (e.g., they could permit for two ad breaks today, four ad breaks tomorrow, and three ad breaks next week), whereby using the actual number of ad breaks per video as observed on a hypothetical day of data collection could entail measurement error. The option to have more ad breaks, however, remains constant over time. Finally, studying the effect of an increase in the feasible number of ad breaks is likely to yield more interesting policy implications. As argued, an increase in the feasible number of ad breaks per video corresponds to repealing restrictions on the YouTubers' advertising quantity, and the insights from this analysis are likely to be more generalizable to other settings than studying the effect of a specific number of additional ad breaks per video.

---

[27]Besides, data on the actual number of ad breaks per video is unavailable, as further explained in Section 4.1.

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 $Pr(D_i = 1)$.

Who comprises the complying population? Compliers are YouTubers who pursue commercial interests, but do not push them excessively (to a potentially unreasonable degree); moreover, the complying population excludes superstars who were aware of the ten minutes trick before Nov 2015. Thus, complying YouTubers resemble conventional media outlets in terms of their characteristics, whereby the LATE from this paper can be informative about the effect of lifting advertising restrictions in media markets beyond YouTube. See Section 10 for further discussions on the external validity of my analysis.

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. Additional requirements for my particular application are discussed in Appendix B.

Instrument relevance   First, the parameter π in the first stage equation (3) must be non-zero, i.e., $close_i$ must be correlated to D_i. It is plausible that the closer a YouTuber's position to the ten minutes threshold before Nov 2015, the easier it is to produce videos that are ten minutes or longer after Nov 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 ($F > 140$).

Exclusion restriction   Second, $close_i$ must operate through the single, known channel D_i. In other words, the instrument must be exogenous to the dependent variable $Mainstream_{vit}$. This is a plausible assumption, too. A YouTuber's median video duration before Nov 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 mainstream 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 $Mainstream_{vit}$ relative to a base month just before the YouTubers could increase their advertising quantity and adapt their content. The event study regression equation is given by

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

The interpretation of this approach is analogous to checking the validity of a parallel trends assumption. While the indirect impact of $close_i$ on $Mainstream_{vit}$ 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 $Mainstream_{vit}$ 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 Nov 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 Nov 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.[28] Third, YouTubers do not benefit from higher values of $close_i$ before Nov 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_{it} in equation (3)) controls for YouTuber specific linear time trends; in Appendix D.3, I also include higher order polynomials of t_{it} into equation (3). Second, while commercial interests are a

---

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

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*.

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 ($10,599$ YouTubers). 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 advertising quantity decreases the probability to duplicate mainstream content by about twenty percentage points. The effect size is considerable: it corresponds to 40% of a standard deviation in the dependent variable $Mainstream_{vit}$ 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 if not taken into account.

The first stage diagnostics in columns 4 to 6 support the validity of my empirical strategy. Having been closer to the ten minutes threshold before Nov 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 mainstream 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 mainstream 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 – i.e., lifting the limitations on advertising quantity – causes a considerable reduction in the YouTubers' probability to duplicate mainstream content. 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, comes along with the incentive to differentiate from competing YouTubers, leading to a lower probability to duplicate mainstream content. A detailed discussion of a plausible economic mechanism follows in Section 7. Appendix C.5 shows that the main results are robust to alternative measures of mainstream content.

6.2. Mechanics

6.2.1. Number of video uploads and keywords

This section shows that the results from Section 6.1 are mostly induced by additional video uploads by YouTubers who could increase their feasible advertising quantity, while the number of keywords per video remains nearly unchanged. To this end, I replace the dependent variable from equation (2) with five alternative variables – the number of video uploads per YouTuber per month, the absolute number of videos with mainstream content per YouTuber per month, the absolute number of keywords per video, the absolute number of keywords per YouTuber per month, and the absolute number of unique keywords per YouTuber per month – and re-run the analysis from above.

Table 4 shows the results. The estimates in columns 1 and 2 are positive; while the potentially biased OLS estimate in column 1 is close to zero and not statistically significant, the IV estimate in column 2 indicates that YouTubers who could increase their feasible ad quantity after Nov 2015 uploaded on average two additional videos per month. This corresponds to about 39.5% of a standard deviation in the dependent variable and to about 47% of its baseline value. However, the results from above are not solely driven by YouTubers uploading additional videos. Column 4 demonstrates that the absolute number of videos with mainstream content uploaded by YouTubers in the treatment relative to the control group decreased by about 0.15 videos per month, which corresponds to 30% of

a standard deviation in the dependent variable and about 35% of its baseline value. Thus, the decrease in probability to duplicate mainstream content as illustrated in Section 6.1 is driven by an increase in the absolute number of video uploads per month and a slight simultaneous decrease in the absolute number of videos covering mainstream content.

In principle, a decrease in the probability to duplicate mainstream content as specified in my application (see Section 4.2) could also be driven by a decrease in the number of keywords per video. Columns 5 and 6, however, show that this is not the case. Both estimates are small and positive; while the IV estimate is not statistically significant, the potentially biased OLS estimate is significant at the 1%-level.

Finally, I explore the number of absolute and the number of unique keywords per YouTuber per month in columns 7 to 10. Consistent with the estimates from columns 2 and 6 (i.e., more videos per month and a constant number of keywords per video), the absolute number of keywords per YouTuber per month increases (columns 7 and 8). Columns 9 and 10 demonstrate that the number of unique keywords per YouTuber per month has increased as well: According to the IV estimate in column 10, YouTubers who could increase the feasible ad quantity per video increased their number of unique keywords per month by about 9.8; the estimate is statistically significant at the 1%-level and corresponds to about 29% of a standard deviation as well as the baseline of the dependent variable. This result shows that the YouTubers not only reduce their duplication of mainstream content, but also diversify the video content that they offer, as further explored below.

6.2.2. 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 YouTubers' probability to duplicate the *most* mainstream 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 mainstream 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 1st to 10th, the 10th to 25th, the 25th to 50th, the 50th to 75th, and the 75th to 100th percentile. Then, I replace the dependent variable $Mainstream_{vit}$ in equation (2) with each of these dummies and estimate equations (2) and (3) by 2SLS.

The results in Table 5 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 1st to 10th 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 sixteen keywords on average and that this number is constant over time (see Section 6.2.1). Hence, a video can be given keywords from several parts of the distribution of most-viewed keywords. Bearing this mind, the estimates in Table 5 demonstrate that the YouTubers who could increase the feasible number of ad breaks per video do not move from exclusively uploading mainstream content to uploading non-mainstream content only. Rather, they change the "mixture" of topics in a video: they abandon covering mainstream topics – the more mainstream, the stronger the effect – and cover more of the less mainstream topics instead. Thereby, the probability to upload very mainstream content decreases, while the probability to upload not so mainstream content remains unchanged or increases only slightly. Indeed, when I count each video's number of "affiliations" to the categories displayed in Table 5 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 Nov 2015 than before (column 6 in Table 5).

6.3. Effect heterogeneity

One particular strength of my dataset is its size, which allows me to conduct several 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

Adapting 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 duplicate mainstream 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.[29] Note that reverse causality prohibits including the subscriber count as an interaction term. If, for instance, YouTubers who upload much mainstream content have a larger audience, I would overestimate the effect of a YouTuber's subscriber count.

Tables 6 and 7 show the results from regressing equations (2) and (3) on the two sub-samples. The potentially biased OLS estimates in columns 1 to 3 are close to zero and

---

[29]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).

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, 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 6, but close to zero for the YouTubers with fewer than $1,000$ subscribers in Table 7. 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, which leads to further effect heterogeneity. 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 8 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 largest 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-mainstream 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. Validity checks

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 $Mainstream_{vit}$ before Nov 2015, but a clear negative and increasingly strong effect after Nov 2015. See Appendix A.1 for further validity checks, including a series of placebo regressions and an analysis of non-advertising Youtubers.

7. Competition between YouTubers

This section studies a plausible economic mechanism driving the results from Section 6. In particular, I show that mainstream content is provided by many competing YouTubers. Thus, viewers could easily switch to a competitor if a YouTuber increased her advertising quantity. Switching is less likely, however, if the YouTuber differentiates her content from the mainstream and thereby softens competition in the ad "price." I support the argument with findings from the YouTubers' audience: if the feasible ad quantity goes up, viewer fluctuation goes down, which is in line with the idea that YouTubers who decrease the probability to duplicate mainstream content gain market power in a niche.

7.1. Definition of competitive content

First, I construct a measure of "competitive content", i.e., a measure for the most-supplied content on YouTube. The procedure is analogous to the definition of "mainstream 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 mainstream 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 mainstream 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 17 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.[30] 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

$$Competitive_{vit} = \beta'D_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \epsilon_{vit}, \tag{6}$$

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 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 (6) 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 (6) by 2SLS.

Finally, the reduced form of equations (3) and (6) is given by

$$Competitive_{vit} = \gamma'close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + v_{vit}, \tag{7}$$

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

The estimates for β' in Table 9 are similar to the estimates for β in Table 3. Columns 1 to 3 in Table 9 show the results from a potentially biased OLS estimation of equation (6). 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 (6), 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 (7). Consistent with the results from the 2SLS estimation, the estimates are negative: a one

---

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

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 statistically significant at the 1%-level.

The results in Table 9 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 $Mainstream_{vit}$ and $Competitive_{vit}$ are highly correlated, this is no surprise. Thus, the estimates confirm that YouTubers who increase their ad "price" soften competition over competitive content, which is a plausible economic mechanism that drives the results from Section 6.

7.3. Viewer fluctuation

If YouTubers differentiate their content from the mainstream to prevent their audience from switching to a competitor, viewer fluctuation should go down, i.e., a given number of views should be generated by a smaller circle of viewers. As precise information on a YouTuber's viewership is not available, I obtain all videos' comments (including the commentators' nicknames) from the YouTube Data API, and employ a YouTuber's commentators as a proxy for her viewers.

The commentator fluctuation of YouTuber i is defined as

$$fluctuation_i = \frac{commentators_i}{comments_i}, \tag{8}$$

where $commentators_i$ corresponds to the size of the set of YouTuber i's eventual commentators, and $comments_i$ corresponds to the total number of comments that i receives. If each comment is written by a different user, commentator fluctuation is maximal and $fluctuation_i$ is equal to 1. The more comments are written by the same commentators, the smaller $fluctuation_i$. 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 Nov 2015,

$$\Delta fluctuation_i = fluctuation_{i,post} - fluctuation_{i,pre}, \tag{9}$$

where $fluctuation_{i,post}$ is based on the fifteen months after, and $fluctuation_{i,pre}$ is based on the fifteen months before Nov 2015.[31] If the commentator fluctuation of YouTuber i goes down, $\Delta fluctuation_i < 0$.

To demonstrate that an increase in the feasible advertising quantity and the corresponding decrease in the YouTubers' probability to upload mainstream content diminishes commentator fluctuation, I use $\Delta fluctuation_i$ as dependent variable in

$$\Delta fluctuation_i = \rho_0 + \rho_1 D_i + \epsilon_i, \tag{10}$$

---

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

where ρ_1 denotes the average effect of having the option to increase the number of ad breaks per video on the change in commentator fluctuation of YouTuber i. To account for endogeneity in D_i, I use

$$D_i = \psi_0 + \psi_1 close_i + e_i, \qquad (11)$$

as a first stage and estimate equations (10) and (11) 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 – e.g., 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 Nov 2015 (see Appendix C.6.2 for alternative thresholds).

Table 10 shows the results. Column 1 presents the potentially biased OLS estimate for ρ_1. The estimate is negative and statistically significant at the 1%-level, i.e., commentator fluctuation decreases if YouTubers have the option to increase the feasible number of ad breaks per video. The 2SLS estimate in column 2 is negative, too, statistically significant at the 5%-level, and more than three times larger in absolute value than the OLS estimate. The effect size corresponds to 18% of a standard deviation in the dependent variable $\Delta fluctuation_i$. The reduced form estimate in column 3 is consistent with the 2SLS estimate in column 2.

8. Differentiation in the aggregate

Up to this point, I have studied how an increase in the feasible advertising quantity 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 Nov 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 accordingly.

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 Nov 2015. As I observe 34 months before Nov 2015, but only 15 months afterwards (including Nov 2015), I limit the analysis to the 15 most recent months before Nov 2015.

In the subsample, there exist $607,358$ unique keywords before, and $875,503$ unique keywords after Nov 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 Nov 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 mainstream or competitive content. The difference in the results could stem from entry: by construction, the population includes all YouTubers who entered the platform after Nov 2015, which may further increase the number of unique keywords that exist after Nov 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 Nov 2015.[32] Again, I restrict the analysis to the 15 months before Nov 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.

The Gini coefficient for the subsample is high and remains nearly unchanged: it is equal to 0.800 before, and equal to 0.806 after Nov 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 Nov 2015. The result does not contradict the findings from Section 7, though. My measures for mainstream 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 Nov 2015 (see Section 8.1). If many of those

---

[32]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 cannot use absolute measures of concentration such as the Herfindahl index, because the number of keywords before and after Nov 2015 is different.

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 Nov 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

To complement the main analysis, this section studies the effect of an increase in the feasible advertising quantity on video quality. Two predictions compete. On the one hand, quality may decrease, since YouTubers soften competition in terms of quality when they differentiate their content from their competitors – an observation that dates back to the seminal paper by Shaked and Sutton (1982).[33] Besides, 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). On the other hand, an increased number of ad breaks per video implies that each viewer is c.p. more valuable than before, whereby the incentive to provide high quality content goes up. Moreover, the YouTubers may want to counterbalance their viewers' additional ad nuisance costs.

9.1. Likes and dislikes

An intuitive way to measure video quality is to consider its number of likes and dislikes, e.g., in terms of a video's fraction of positive ratings:

$$\left(\frac{Likes}{Likes + Dislikes}\right)_{vit}. \tag{12}$$

To examine the effect of an increase in the feasible advertising quantity on video quality, I replace the dependent variable in equation (2) with expression (12) and estimate equations (2) and (3) by 2SLS.

Table 11 shows the results.[34] The potentially biased OLS estimates of equation (2) (columns 1 to 3) are close to zero and not statistically significant. In contrast to that,

---

[33]See, e.g., Bourreau (2003); Armstrong and Weeds (2007); Weeds (2013) for applications on media markets.

[34]Note that $77,066$ videos have not received any likes or dislikes.

the 2SLS estimates in columns 4 to 6 are negative and statistically significant at the 1%-level. According to these estimates, an increase in the feasible advertising quantity leads to a four percentage point reduction in the fraction of positive ratings. The effect size corresponds to around 25% of a standard deviation in the dependent variable and to 4.4% of its baseline value 0.91. The reduced form estimates in columns 7 to 9 are in line with these findings. Thus, the viewers' satisfaction with video consumption seems to diminish when the feasible advertising quantity goes up and YouTubers differentiate their content from their competitors. It is, however, unclear if this effect is driven by changes in video quality as such or if must at least partially be attributed to the viewers' ad aversion. It is, for instance, possible that viewers dislike a video to express their irritation from additional ad breaks even though video quality has gone up.

9.2. Views

If video quality enhances, it is also possible that the number of viewers goes up or that viewers watch a video repeatedly. Thus, I use a video's log number of views, $log(Views)_{vit}$, as a second measure of its quality.[35]

As in Section 9.1, I replace the dependent variable in equation (2) with the quality metric and estimate equations (2) and (3) by 2SLS. Table 12 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 in columns 7 to 9 match 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, 2019, 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.

---

[35]YouTube counts a view if the video is watched for at least thirty seconds; if the video is shorter than that, the viewer must watch it entirely to be counted. See `www.tubics.com/blog/what-counts-as-a-view-on-youtube/` (May 2019).

10. Conclusion

This paper demonstrates that an increase in the feasible advertising quantity leads to an increase in content differentiation between several thousand YouTubers. In particular, I show that the option to raise the number of ad breaks per video reduces the YouTubers' probability to duplicate mainstream content by about twenty percentage points, because YouTubers differentiate their content from their competitors to soften competition in the ad "price." The results support models that acknowledge a conceptual equivalence of advertising and subscription prices as determinants of content differentiation in media markets, and provide empirical evidence of the interplay of price competition and horizontal product differentiation in general.

Regarding its tremendous reach, analyses of YouTube are interesting and relevant by themselves. Moreover, its institutional features allow me to identify causal effects. The empirical findings are, however, likely to be valid beyond the specific context of this paper. In particular, they are likely to hold for other media markets where consumers perceive advertising as a nuisance and where many competing media outlets co-exist: e.g., television and radio broadcasters as well as (online) newspapers might further differentiate their content from their competitors if upper bounds on advertising quantity are lifted. Although YouTubers may be different from conventional media outlets – e.g., channels are usually driven by just one or a small group of individuals – many YouTubers run their channel like a proper business and the complying population shows a reasonable amount of commercial interest, whereby the insights from this paper are applicable beyond YouTube.

The paper advances persistent debates on the effect of advertising on content differentiation. In particular, showing that advertising does *not* lead to the duplication of mainstream content entails two implications for present policies. First, advertising quantities are often restricted in an attempt to protect consumers.[36] 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 four respects that open up perspectives for further research. First, although I present competition in the ad "price" as a plausible mechanism for my main results and rule out a YouTuber learning effect, I cannot exclude the possibility that

---

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

there are other potential mechanisms. For instance, YouTubers might not only soften competition to other YouTubers and acquire a more stable audience when they upload less mainstream content, but the characteristics of their viewers may change, too. Viewers of less mainstream 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 fixed and, in particular, 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. A downward sloping (inverse) demand curve on behalf of the advertisers, however, may have the opposite effect. If the price per ad declines in quantity, the effect of an increase in the feasible advertising quantity on content differentiation would not be reversed, but may be reduced.

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. Validity checks

This section presents additional validity checks for my empirical strategy.

A.1. Placebo regressions

First, I conduct a series of placebo regressions. To this end, I augment the reduced form equation (4) to

$$Mainstream_{vit} = \gamma^{Placebo} close_i * fakepost_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + v_{vit} \mid t \leq 34, \quad (13)$$

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 30 placebo regressions in sum. If $close_i$ has no direct effect on $Mainstream_{vit}$, 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 $Mainstream_{vit}$ before Nov 2015.

Of 30 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.2. Non-advertising YouTubers

The non-advertising YouTubers, whom I do not consider in the main analysis, allow me to conduct additional validity checks. The non-advertising YouTubers' content choices are not driven by commercial considerations. As a consequence, their probability to duplicate mainstream content cannot be affected by the launch of the new ad break tool in Nov 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 results in Table 13 support these considerations: the IV estimates for β are small and statistically insignificant for ten out of twelve regressions.

Figure 18 provides an additional plausibility check of the exclusion restriction. If the only way the instrument $close_i$ affects the dependent variable $Mainstream_{vit}$ 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 18 demonstrates that this is the case.

B. Additional requirements

In addition to the validity of the instrument, two further requirements must be fulfilled in my application. First, to be consistent with the idea of the identification strategy, the effect of an increase in the feasible number of ad breaks per video on the probability to upload mainstream 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 mainstream content. These additional requirements are now discussed.

B.1. 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 duplicate mainstream 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. To confirm that this aggregate effect is driven by videos that are ten minutes or longer, I augment the first stage regression equation (3) to

$$I(\geq 10)_{vit} = \alpha close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + v_{vit} | Mainstream_{vit}, \qquad (14)$$

where $I(\geq 10)_{vit}$ indicates if video v of YouTuber i in month t is ten minutes or longer. Then, I estimate equation (14) by OLS for mainstream and for non-mainstream videos 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 mainstream 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 YouTubers' probability to duplicate mainstream 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-mainstream content*, because the probability that a video is ten minutes or longer within this subsample increases after Nov 2015. In contrast to that, the OLS estimate for α should be close to zero and not statistically significant when I condition on mainstream content. Note that reverse causality concerns prohibit an interaction of the term $close_i * post_t$ in equation (14) with a dummy that indicates mainstream 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 duplicate mainstream content, the estimate for the triple interaction would be downward biased.

Table 14 displays the results from an OLS regression of equation (14) on the subsample of mainstream videos (columns 1 to 3) and on the subsample of non-mainstream 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 B. 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 mainstream, and positive when considering only non-mainstream content.

B.2. Video duration and mainstream content

When a YouTuber is treated, she does not just benefit from a lift of her ad quantity restrictions, but – by construction – her videos' duration increases, too. Hence, I must also ensure that video duration as such does not affect the dependent variable $Mainstream_{vit}$.

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 Nov 2015 may be manipulated to exploit the ten minutes trick. In contrast to that, the videos' duration *before* Nov 2015 is – similar to the instrument $close_i$ – likely to be as good as randomly assigned. Hence, I run the following regression

$$Mainstream_{vit} = \delta duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \epsilon_{vit} \,|\, t \leq 34 \qquad (15)$$

including only observations from before Nov 2015, where I expect δ to be close to zero and statistically insignificant.

Table 15 shows that the estimate for δ in equation (15) is very small and statistically insignificant. Thus, I find no evidence in my data that video duration as such directly affects $Mainstream_{vit}$.

C. 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 alternative selections of YouTubers, to alternative classifications of the treatment group, and to alternative definitions of the instrument $close_i$ and of the dependent variable $Mainstream_{vit}$. 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.

C.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 16 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 Nov 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 6.4 illustrates that the effect of an increase in the feasible number of ad breaks on the probability to upload mainstream content becomes stronger over time. Thus, excluding the last six months from the analysis results in smaller estimates.

C.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 Nov 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 Nov 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 Nov 2015 is smaller than 7.5, second, a subsample of YouTubers whose 90th percentile of the distribution of video durations (not the median) is smaller than 10.

Tables 17 and 18 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.

C.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 19 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 mainstream 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 mainstream 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 20 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.

C.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 Nov 2015 in the main analysis, $close_i$ corresponds to the 75th and to the 90th 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 Nov 2015.

Table 21 shows the results from a 2SLS estimation of equations (2) and (3) using a YouTuber's 75th percentile (columns 1 to 3), and using a YouTuber's 90th percentile of

the distribution of video durations before Nov 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 75th and the 90th percentiles of the distribution of a YouTuber's video durations before Nov 2015 have less power to predict a YouTuber's treatment status D_i than the median.

C.5. Alternative definitions of mainstream content

Here, I show that the results from Section 6.1 are robust to alternative definitions of mainstream 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 mainstream – 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 mainstream 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 mainstream and all videos given such a keyword are assigned a dummy equal to one. Table 24 provides an overview of how these measures are correlated.

Table 25 shows the results from a 2SLS estimation of equations (2) and (3) using the four alternative definitions of $Mainstream_{vit}$. 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 mainstream 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.

C.6. 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 $Mainstream_{vit}$ 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 7.3.

C.6.1. Alternative definitions of competitive content

Analogous to Appendix C.5, I show that the results from Section **??** 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 26 provides an overview of how these measures are correlated.

Table 27 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.

C.6.2. Alternative cutoffs in the commentator analysis

Here, I show that the results from Section 7.3 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 Nov 2015.

The results in Table 22 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 Nov 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 10. In contrast to that, when I restrict the analysis to YouTubers with at least ten comments before and after Nov 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.

D. 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.

D.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.

D.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 Nov 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 13 may be too large, too. This would, however, strengthen my results: Section A.2 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.

D.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.

D.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 mainstream 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.

D.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.[37] 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 28. In a last step, I review all posts from Table 28 and indicate if a YouTuber's monetization options or her probability to upload mainstream content could be affected. Thirteen events require further investigation; I discuss them chronologically.

D.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,

---

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

however, extremely limited: not even 100 YouTubers worldwide participated.[38] 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.

D.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 $close_i$.

D.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.

---

[38]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.

D.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 mainstream 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 mainstream content after Nov 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_{it} controls for a YouTuber's average change in the probability to upload mainstream (or competitive) content over time. Columns 1 and 4 in Table 29 replicate the 2SLS results from Tables 3 and 9 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_{it}. On the other hand, the estimates for t_{it}, though negative, are extremely small. A YouTuber's probability to upload mainstream 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_{it}^2 and t_{it}^3 does not affect the estimates for β and β', either (columns 2, 3, 5, and 6 of Table 29). 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 mainstream or competitive content increases in the beginning, but decreases from around her 160th 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," *The 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 (2015) "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*.

Baker, Matthew and Lisa M George (2019) "The news hour: Welfare estimation in the market for local tv news," *Available at SSRN 2694687*.

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

Berry, Steven T and Joel Waldfogel (2001) "Do mergers increase product variety? Evidence from radio broadcasting," *The Quarterly Journal of Economics*, **116** (3), 1009–1025.

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.

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," *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.

Davcik, Nebojsa S and Piyush Sharma (2015) "Impact of product differentiation, marketing investments and brand equity on pricing strategies," *European Journal of Marketing*.

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

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 (1984) "The principle of minimum differentiation revisited," *European Economic Review*, **24** (3), 345–368.

——— (1986) "Minimal and maximal product differentiation in Hotelling's 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.

Gentzkow, Matthew, Jesse M Shapiro, and Daniel F Stone (2015) "Media bias in the marketplace: Theory," in *Handbook of Media Economics*, **1**: Elsevier, 623–645.

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

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

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," *The 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.

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.

Netz, Janet S and Beck A Taylor (2002) "Maximum or minimum differentiation? Location patterns of retail outlets," *Review of Economics and Statistics*, **84** (1), 162–175.

Numan, Wieteke and Marleen Willekens (2012) "An empirical test of spatial competition in the audit market," *Journal of Accounting and Economics*, **53** (1-2), 450–465.

Petrova, Maria (2011) "Newspapers and parties: How advertising revenues created an independent press," *American Political Science Review*, 790–808.

Puglisi, Riccardo and James M Snyder (2015) "Empirical studies of media bias," in *Handbook of Media Economics*, **1**: Elsevier, 647–667.

Salganik, Matthew J (2019) *Bit by bit: Social research in the digital age*: Princeton University Press.

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.

Shaked, Avner and John Sutton (1982) "Relaxing price competition through product differentiation," *The Review of Economic Studies*, 3–13.

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.

Sweeting, Andrew (2010) "The effects of mergers on product positioning: evidence from the music radio industry," *The RAND Journal of Economics*, **41** (2), 372–397.

Tirole, Jean (1988) *The Theory of Industrial Organization*: MIT press.

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.
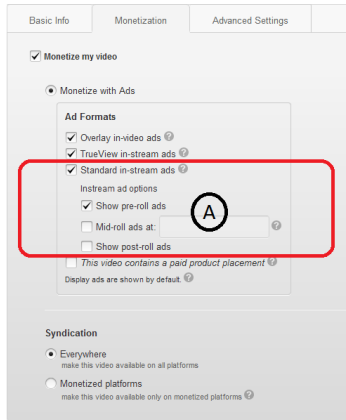
Figure 1: Old ad break tool (before Nov 2015).
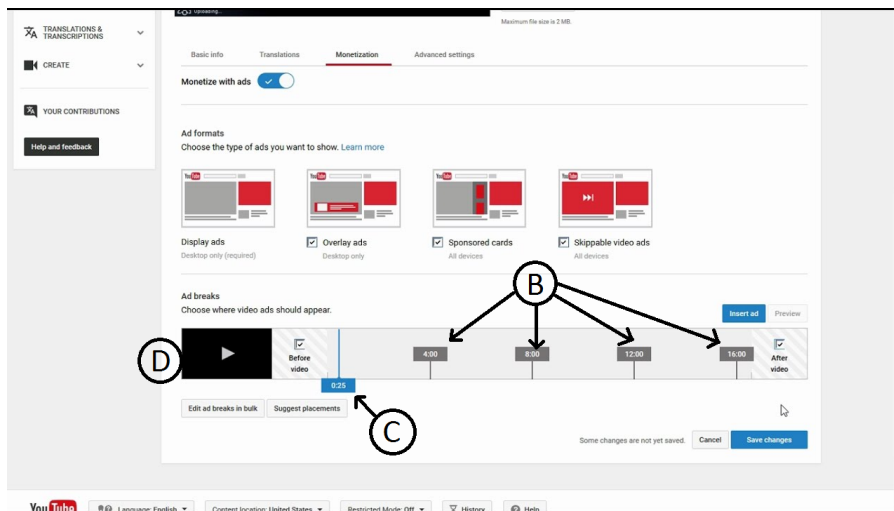
Figure 2: New ad break tool (after Nov 2015).
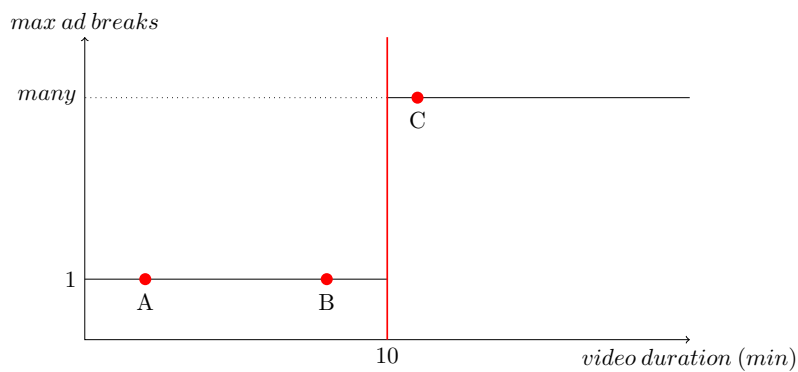
Figure 3: Stylized example of the identification strategy.
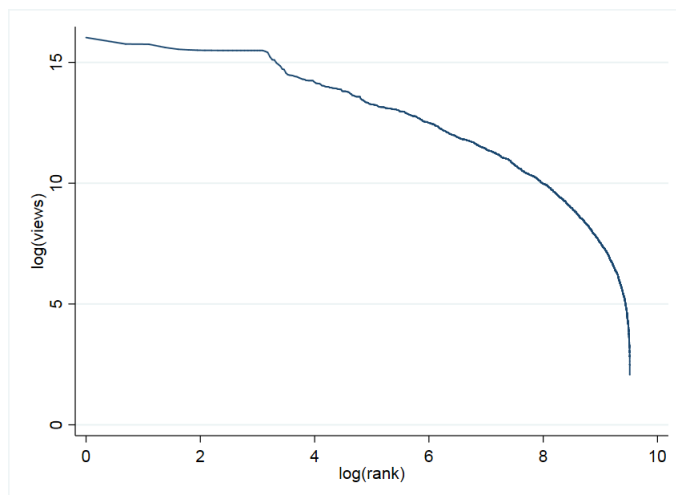
Figure 4: Log-log plot of the number of views a keyword attracts and its associated rank in the category "Science & Technology" in April 2015.
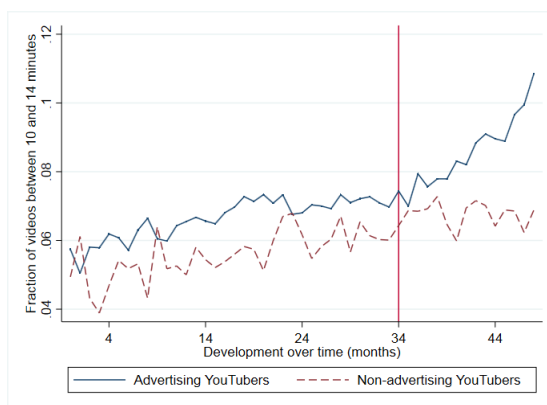
Figure 5: Trends in advertising vs. non-advertising YouTubers' share of videos between ten and fourteen minutes. The vertical line depicts Nov 2015.
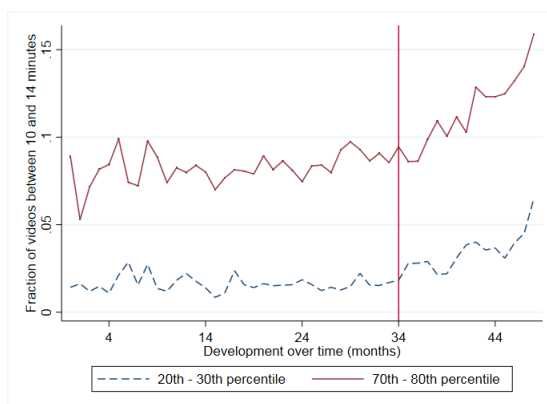
Figure 6: Trends in advertising YouTubers' share of videos between ten and fourteen minutes. The vertical line depicts Nov 2015.
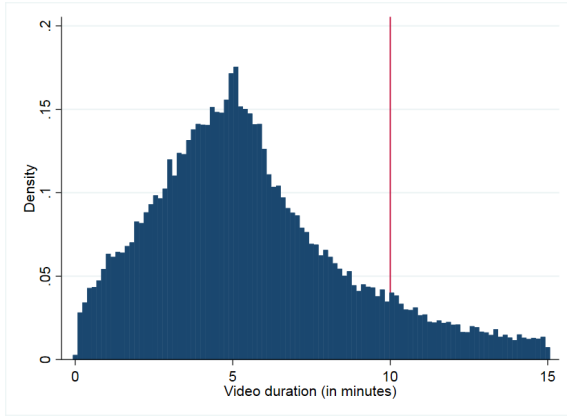
Figure 7: Histogram of the distribution of video durations *before* Nov 2015 for YouTubers from the 70th to 80th percentile in median video duration before Nov 2015. The vertical line depicts the ten minutes threshold.
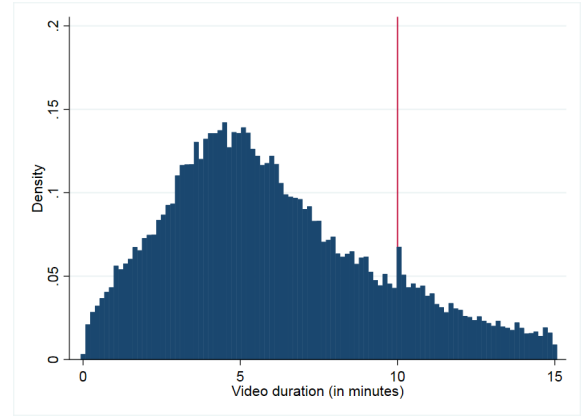
Figure 8: Histogram of the distribution of video durations *after* Nov 2015 for YouTubers from the 70th to 80th percentile in median video duration after Nov 2015. The vertical line depicts the ten minutes threshold.
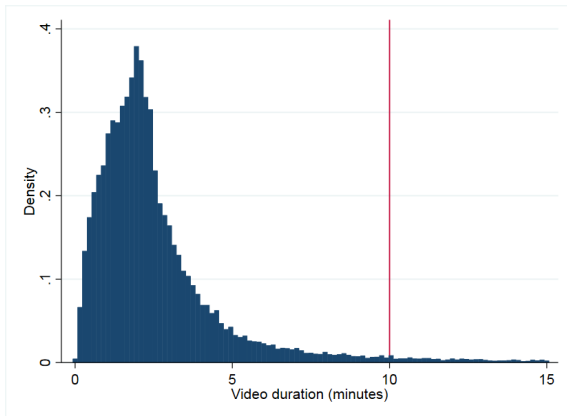
Figure 9: Histogram of the distribution of video durations *before* Nov 2015 for YouTubers from the 20th to 30th percentile in median video duration before Nov 2015. The vertical line depicts the ten minutes threshold.
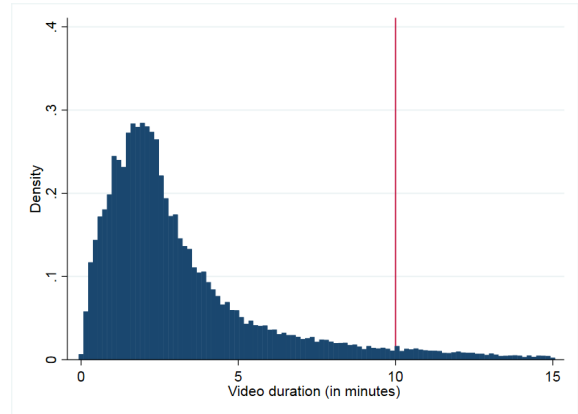
Figure 10: Histogram of the distribution of video durations *after* Nov 2015 for YouTubers from the 20th to 30th percentile in median video duration after Nov 2015. The vertical line depicts the ten minutes threshold.
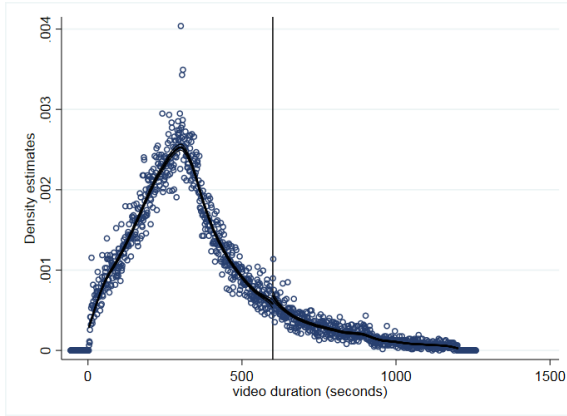
Figure 11: McCrary test of the distribution of video durations *before* Nov 2015 for YouTubers from the 70th to 80th percentile in median video duration before Nov 2015. The vertical line depicts the ten minutes threshold.
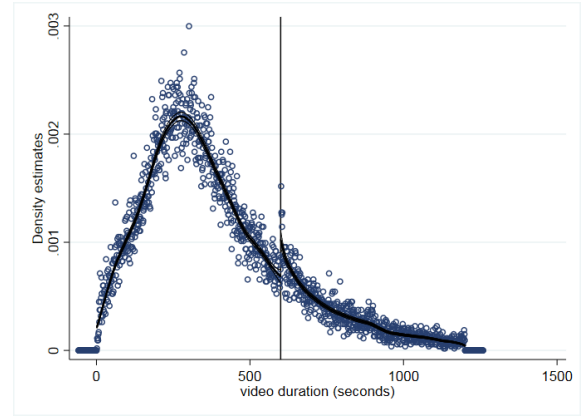
Figure 12: McCrary test of the distribution of video durations *after* Nov 2015 for YouTubers from the 70th to 80th percentile in median video duration before Nov 2015. The vertical line depicts the ten minutes threshold.
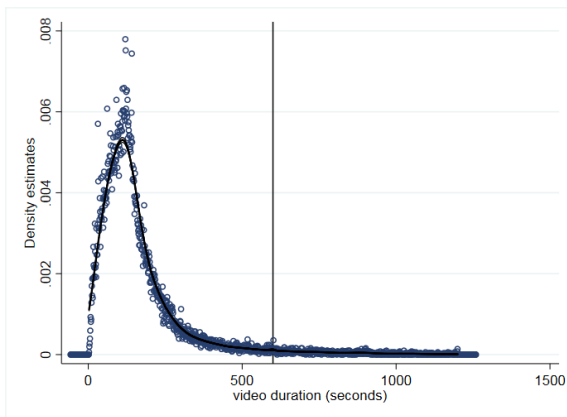
Figure 13: McCrary test of the distribution of video durations *before* Nov 2015 for YouTubers from the 20th to 30th percentile in median video duration before Nov 2015. The vertical line depicts the ten minutes threshold.
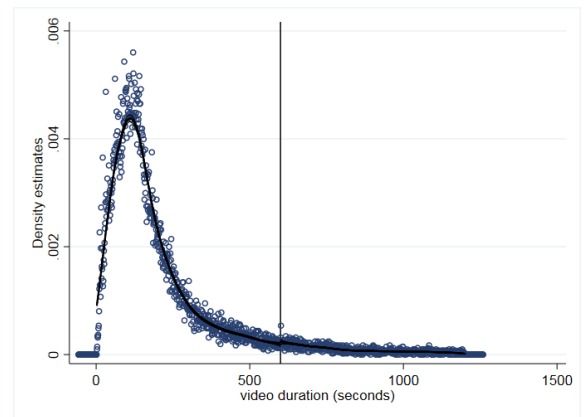
Figure 14: McCrary test of the distribution of video durations *after* Nov 2015 for YouTubers from the 20th to 30th percentile in median video duration before Nov 2015. The vertical line depicts the ten minutes threshold.
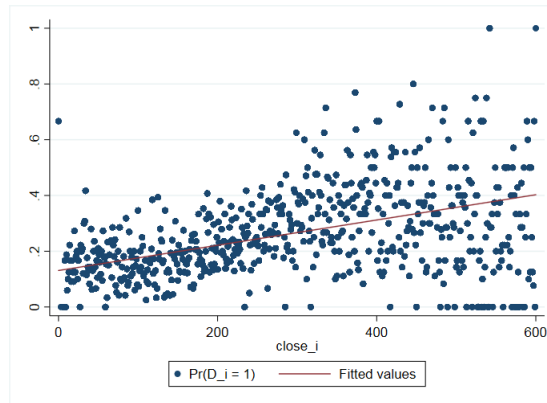
Figure 15: Plot of all values of $close_i$ on the associated average probability to be treated, $Pr(D_i = 1)$.
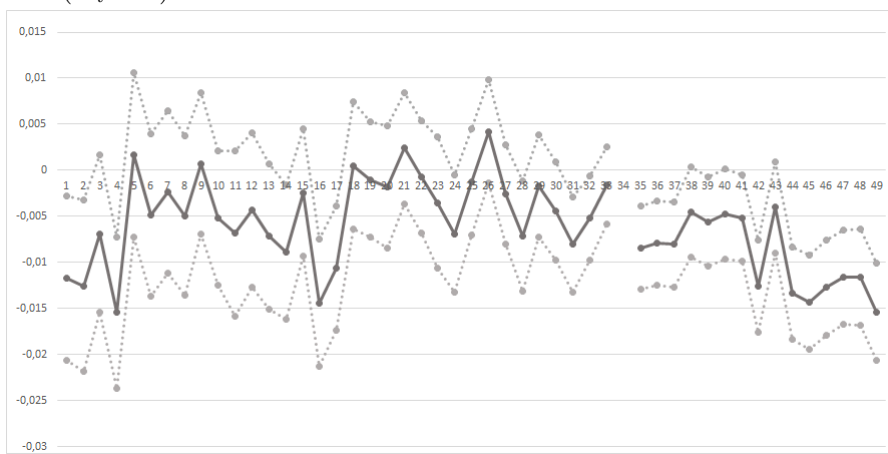
Figure 16: Event study mainstream content (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t$ from equation (5). The dashed line depicts a 95%-confidence interval.
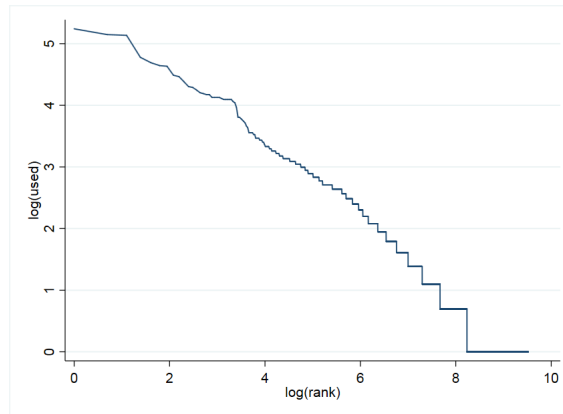
Figure 17: Log-log plot of the number of usages of a keyword and their associated rank in the category "Science & Technology" in April 2015.
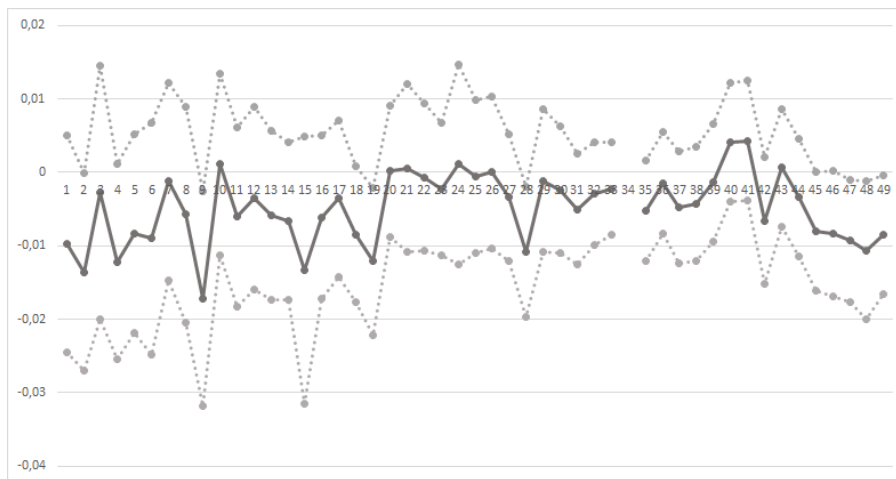
Figure 18: Event study mainstream content (non-advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t$ from equation (5). The dashed line depicts a 95%-confidence interval.

Table 1: Summary statistics

| Variable	Mean	Std. Dev.	Min.	Max.	N
$Mainstream_{vit}$	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
$Let's Play_{vit}$	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
$Uploads_{it}$	4.337	5.269	1	249	322,200
$MainstreamMonth_{it}$	0.401	0.49	0	1	322,200
$NoKeywords_{vit}$	14.859	8.736	1	67	1,397,267
$AbsKeywords_{it}$	64.438	90.626	1	4482	322,200
$UniKeywords_{it}$	32.59	32.324	1	1059	322,200
$I(1^{st}to10^{th})_{vit}$	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 $Mainstream_{vit}$, $Competitive_{vit}$, $Advertising_i$, $post_t$, 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 Nov 2015 (Figure 11)	0.1654*** (0.0640)
70th to 80th percentile post Nov 2015 (Figure 12)	0.4049*** (0.0604)
20th to 30th percentile pre Nov 2015 (Figure 13)	0.0035 (0.1609)
20th to 30th percentile post Nov 2015 (Figure 14)	0.2659** (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 (0.0078)	0.006 (0.0077)	0.006 (0.0077)	-0.209*** (0.0505)	-0.200*** (0.0491)	-0.192*** (0.0480)			
$close_i * post_t$							-0.006*** (0.0013)	-0.006*** (0.0013)	-0.006*** (0.0013)
First stage				0.0286*** (0.0023)	0.0286*** (0.0023)	0.0290*** (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	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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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: Mechanics

	Uploads		No. Mainstream		Keyw./video		Abs. keyw./month		Unique keyw./month	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$D_i * post_t$	0.398	2.071***	0.010	-0.148***	0.413***	0.321	7.973***	39.290***	2.430***	9.795***
	(0.082)	(0.575)	(0.006)	(0.042)	(0.136)	(0.800)	(1.601)	(10.685)	(0.584)	((3.763)
First stage		0.028***		0.028***		0.029***		0.028***		0.028***
		(0.002)		(0.002)		(0.002)		(0.002)		(0.002)
F-test of excluded instruments		202.12		202.12		151.32		202.12		202.12
Time FE	X	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X	X
Controls					X	X				
YouTuber Time Trend	X	X	X	X	X	X	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
Videos	241,905	241,905	241,905	241,905	1,067,542	1,067,542	241,905	241,905	241,905	241,905

Notes: Robust standard errors in parentheses. The dependent variable in columns 1 and 2 is the number of video uploads of YouTuber i in month t. The dependent variable in columns 3 and 4 is the absolute number of videos that cover mainstream content of YouTuber i in month t. The dependent variable in columns 5 and 6 is the number of keywords of video v of YouTuber i in month t. The dependent variable in columns 7 and 8 is the absolute number of keywords that YouTuber i uses in month t. The dependent variable in columns 9 and 10 is the number of unique keywords that YouTuber i uses in month t. 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 5: Differentiation along the tail

	1st to 10th percentile	10th to 25th percentile	25th to 50th percentile	50th to 75th percentile	75th to 100th percentile	Sum of affiliations
	(1)	(2)	(3)	(4)	(5)	(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	168.04	169.44	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	10,590	10,589	10,589
Videos	1,064,248	1,033,666	1,033,666	1,031,051	1,028,446	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 1st to 10th 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 10th to 25th 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 25th to 50th 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 50th to 75th 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 75th to 100th 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 6: 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)	(0.0097)	(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***	0.026***	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	5,183	5,183	5,183	5,183	5,183	5,183	5,183	5,183	5,183
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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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 7: Main regression, subscribers $< 1,000$

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	0.015	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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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 8: 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	-0.658	-0.243**	-0.548	-0.255**	-0.236	-0.100	-0.651	-0.227
	(0.3405)	(0.0953)	(12.376)	(0.179)	(0.134)	(0.139)	(0.852)	(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 $Mainstream_{vit}$ 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 9: 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	-0.004	-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
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 10: 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 Nov 2015 for YouTuber i. The estimates are based on the advertising YouTubers only. Only YouTubers who received more than 25 comments before and after Nov 2015 are included in the analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: 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)	(0.0015)	(0.0108)	(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 12: 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				144.26	143.98	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	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081

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 13: Main results non-advertising YouTubers

	Main results (1)	Uploads (2)	No. Mainstream (3)	Keywords/ Video (4)	Abs. keyw. Month (5)	Unique keyw. Month (6)	1st to 10th percentile (7)	10th to 25th percentile (8)	25th to 50th percentile (9)	50th to 75th percentile (10)	75th to 100th percentile (11)	Sum of affiliations (12)
$D_i * post_t$	-0.012	-0.543	0.017	-0.298	0.693	-6.856	-0.169**	-0.146*	-0.027	-0.027	0.019	-0.305
	(0.089)	(1.048)	(0.075)	(1.510)	(17.14)	(5.294)	(0.077)	(0.082)	(0.075)	(0.079)	(0.068)	(0.197)
First stage	0.024***	0.024***	0.024***	0.024***	0.024***	0.024***	0.024***	0.025***	0.025***	0.025***	0.025***	0.025***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
F-test of excluded instruments	47.70	75.15	75.15	47.70	75.15	75.15	48.25	53.44	53.44	54.09	55.43	55.43
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	X	X	X	X	X	X	X	X
YouTubers	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278
Videos	329,725	80,295	80,295	329,725	80,295	80,295	328,544	314,980	314,980	313,665	312,398	312,398

Notes: Robust standard errors in parentheses. The dependent variable in column 1 is $Mainstream_{vit}$, 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 dependent variable in column 2 is the number of video uploads of YouTuber i in month t. The dependent variable in column 3 is the absolute number of videos that cover mainstream content of YouTuber i in month t. The dependent variable in column 4 is the number of keywords of video v of YouTuber i in month t. The dependent variable in column 5 is the absolute number of keywords that YouTuber i uses in month t. The dependent variable in column 6 is the number of unique keywords that YouTuber i uses in month t. In column 7, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 1st to 10th percentile of the distribution of most-viewed keywords. In column 8, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 10th to 25th percentile of the distribution of most-viewed keywords. In column 9, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 25th to 50th percentile of the distribution of most-viewed keywords. In column 10, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 50th to 75th percentile of the distribution of most-viewed keywords. In column 11, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 75th to 100th percentile of the distribution of most-viewed keywords. In column 12, the dependent variable is the sum of most-viewed keywords. In column 12, the dependent variable is the sum of a video's percentile indicators. 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 14: Mainstream content – Evidence from the video level

	Mainstream (1)	Mainstream (2)	Mainstream (3)	Non-main. (4)	Non-main. (5)	Non-main. (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 mainstream. The estimates in Columns 4 to 6 are based on videos classified as non-mainstream. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Video duration and mainstream content

	OLS (1)	OLS (2)	OLS (3)
video duration	0.0000760	0.0000125	0.0000119
	(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

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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 $t \geq 35$. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: 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	0.005	0.005	-0.131***	-0.130***	-0.124***			
	(0.008)	(0.008)	(0.008)	(0.048)	(0.047)	(0.046)			
$close_i * post_t$							-0.004***	-0.004***	-0.004***
							(0.001)	(0.001)	(0.001)
First stage				0.029***	0.029***	0.029***			
				(0.002)	(0.002)	(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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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 17: 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	0.005	0.005	-0.246***	-0.240***	-0.234***			
	(0.008)	(0.008)	(0.008)	(0.050)	(0.049)	(0.048)			
$close_i * post_t$							-0.010***	-0.010***	-0.009***
							(0.002)	(0.002)	(0.002)
First stage				0.040***	0.040***	0.040***			
				(0.003)	(0.003)	(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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers whose median video duration before Nov 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 18: Alternative selection of YouTubers – 90th 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 (0.011)	0.005 (0.010)	0.004 (0.011)	-0.213*** (0.057)	-0.234*** (0.056)	-0.234*** (0.057)			
$close_i * post_t$							-0.011*** (0.003)	-0.013*** (0.003)	-0.013*** (0.003)
First stage				0.054*** (0.005)	0.053*** (0.005)	0.053*** (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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers whose 90th percentile of the distribution of video durations before Nov 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 19: 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*** (0.062)	-0.234*** (0.060)	-0.232*** (0.060)	-0.363*** (0.090)	-0.347*** (0.087)	-0.329*** (0.084)
First stage	0.024 *** (0.003)	0.024*** (0.002)	0.024*** (0.003)	0.016*** (0.001)	0.016*** (0.002)	0.017*** (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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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 Nov 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 Nov 2015 are classified as treated. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 20: 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	-0.001	-0.0008	-0.151***	-0.145***	-0.141***			
	(0.007)	(0.007)	(0.007)	(0.036)	(0.035)	(0.035)			
$close_i * post_t$							-0.006***	-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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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 21: Alternative definitions of the instrument

	75th perc. (1)	75th perc. (2)	75th perc. (3)	90th perc. (4)	90th perc. (5)	90th perc. (6)
$D_i * post_t$	-0.131*	-0.148**	-0.140*	-0.377**	-0.425***	-0.413***
	(0.074)	(0.074)	(0.073)	(0.148)	(0.161)	(0.157)
First stage	0.007***	0.007***	0.007***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.0006)	(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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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 75th percentile in the distribution of video durations of YouTuber i before Nov 2015. In columns 4 to 6, the instrument $close_i$ is defined as the 90th percentile in the distribution of video durations of YouTuber i before Nov 2015. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 22: Alternative cutoffs for the commentator analysis

	> 100 OLS (1)	> 100 2SLS (2)	> 100 Red. Form (3)	> 50 OLS (4)	> 50 2SLS (5)	> 50 Red. Form (6)	> 10 OLS (7)	> 10 2SLS (8)	> 10 Red. Form (9)	> 0 OLS (10)	> 0 2SLS (11)	> 0 Red. Form (12)
D_i	-0.009** (0.004)	-0.066*** (0.024)		-0.01** (0.004)	-0.050** (0.022)		-0.017*** (0.004)	-0.035 (0.025)		-0.035*** (0.005)	0.038 (0.031)	
$close_i$			-0.002*** (0.0007)			-0.002** (0.0007)			-0.0010 (0.0459)			0.001 (0.0008)
First stage		0.031*** (0.0028)			0.031*** (0.0025)			0.028*** (0.0021)			0.028*** (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 Nov 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 Nov 2015 are included in the analysis. In columns 4 to 6, only YouTubers who received more than 50 comments before and after Nov 2015 are included in the analysis. In columns 7 to 9, only YouTubers who received more than 10 comments before or after Nov 2015 are included in the analysis. In columns 10 to 12, all YouTubers who received at least one comment before and after Nov 2015 are included in the analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

65

Table 23: Video duration and likes/(likes+dislikes) and log(Views)

	$\dfrac{Likes}{Likes + Disl}$	$\dfrac{Likes}{Likes + Disl}$	$\dfrac{Likes}{Likes + Disl}$	$log(Views)$	$log(Views)$	$log(Views)$
video	0.000109***	0.000105***	0.000109***	0.0146***	0.0146***	0.0146***
duration	(0.000035)	(0.000035)	(0.000035)	(0.0013)	(0.0013)	(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 24: Correlation measures mainstream 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 mainstream content.

Table 25: Alternative definitions of mainstream 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 $Mainstream_{vit}$ 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 $Mainstream_{vit}$ 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 $Mainstream_{vit}$ 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 $Mainstream_{vit}$ 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

67

Table 26: 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 27: 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 28: 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 video's 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 29: Learning

	Mainstream (1)	Mainstream (2)	Mainstream (3)	Competitive (4)	Competitive (5)	Competitive (6)
$D_i * post_t$	-0.192***	-0.201***	-0.196***	-0.179***	-0.188***	-0.185***
	(0.0480)	(0.0485)	(0.0482)	(0.0456)	(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 $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$