Subscription Revenue and Advertising Strategies on YouTube

Aniket Panjwani
Northwestern University

Heyu Xiong
Case Western Reserve University *

November 23, 2023

Abstract

This paper studies the patterns of substitution between advertising and subscription revenue for media creators on YouTube. YouTube creators exercise enormous discretion in how they monetize their video content. Utilizing a natural experiment on Patreon, a platform that allows YouTubers to solicit recurring payments directly from their viewers, we examine how viewers' willingness-to-pay for content affects the advertisement bundle within YouTube videos. We show that YouTubers respond to the exogenous loss of subscription revenue by increasing affiliate marketing as well as the overall commercial content embedded within their videos. On the other hand, we find no significant effect of subscription revenue on explicitly disclosed sponsored content or the frequency of in-feed ad breaks as determined by video duration. Altogether, our results suggest that changes in subscription revenue may induce changes, not just in the level of advertising, but also in the advertisement mix within media content.

*We are grateful to seminar participants at the ISMS Marketing Science Conference, NBER IT & Digitization Summer Institute, Northeast Ohio Economics Workshop, and the Economics Media Bias Workshop for helpful advice and comments. Panjwani: Northwestern University (email:aniketpanjwani2018@u.northwestern.edu Xiong: Case Western Reserve University, Weatherhead School of Management, (email:hxx226@case.edu).

1 Introduction

The media industry operates as a two-sided market: attracting consumers with content and selling their attention to advertisers. Content providers derive revenue from both sides of the market and historically, advertisers have subsidized consumers (Pattabhiramaiah, Sriram and Sridhar, 2018). In recent years, however, the economics of ad-financed media has been challenged by increased competition for ad-dollars and the proliferation of ad-avoidance technology.[1] At the same time, digital media delivery has also made it easier for publishers and content providers to monetize and sell directly to their end users.[2] As a result of these structural changes, the shift toward subscription-driven business models has become a dominant theme in the media industry.

While the trade-off between subscribers and advertisers features centrally in theoretical models of media (Depken II and Wilson, 2004; Anderson and Gabszewicz, 2006), the broad implications of the transition from ad-sponsored to subscription-based media remain largely unexplored. Little empirical research speaks to either the direct effect of subscription revenue on media content or its downstream effect on the market for advertising.

In this paper, we study how consumers' willingness-to-pay for content affects advertising strategies pursued by media creators and the level of advertising embedded in their products. In particular, we explore how exogenous changes in potential subscription revenue generated by media content induce shifts not just in the quantity of advertising but also in the type of advertising contained in the content.

We investigate these questions in the context of the second most visited website in the world — YouTube (Kerkhof, 2020). YouTube is a media platform that primarily features user-generated content produced by independent content creators (YouTubers). Importantly, YouTubers can choose

[1]For instance, between 2000 and 2017 in the United States, total advertising spending in daily print newspapers declined by over 60% (Angelucci and Cagé, 2019). The large supply of social media platforms on which to serve digital ads has increased competition and reduced ad-prices, forcing media companies to explore new sources of revenue. Simultaneously, the proliferation of ad-avoiding technology like TiVo and Ad-Block has also exerted downward pressure on advertising revenue (Anderson and Gans, 2008).

[2]For example, the New York Times enacted a paywall on its digital content in 2011. As of 2016, the publication has over 1.6 million digital subscribers and the revenue from subscriptions exceeds those coming from advertising. See: https://medium.com/the-graph/advertising-vs-subscription-a4200642842e

to monetize their videos in a variety of ways. For instance, they can solicit payments directly from viewers or pursue advertising-based strategies. Within the broad category of advertising, YouTubers also have considerable latitude in how they incorporate marketing into their content. Common methods of advertising on YouTube include in-feed ad breaks, affiliate marketing, sponsored content, and product reviews and/or placements. Implicitly, YouTubers often face a trade-off between these different monetization schemes. Given the large number of creators who make high-frequency decisions regarding the revenue structure of their media content on the platform, YouTube serves as an excellent empirical laboratory that is uniquely suited to examine the complext interactions between the different monetization methods pursued by media providers.

In order to identify the impact of subscription revenue, we focus on YouTubers who are active on Patreon. Patreon is a subscription membership platform launched in 2013 that helps content creators to manage and collect recurring payments from their viewers and supporters. While YouTubers had some limited means solicit donations directly from viewers prior to Patreon, as a payment processor designed expressly for this purpose, Patreon greatly streamlined the process of collecting subscription revenue and enabled that possibility for a significantly larger share of YouTube creators.

To isolate exogenous variation in Patreon revenue, we utilize a natural experiment on the platform. In October 2015, Patreon's database was hacked and made publicly available on the internet. Once the hack became widely publicized, many patrons canceled their recurring subscriptions on Patreon. This translated to a significant negative shock to the amount of subscription income generated on this platform for YouTubers. We exploit variations in the size of the shock across YouTubers in conjunction with the timing of the hack to analyze how YouTube creators adjusted the advertising content of their products in response to changes in viewers' propensity to subscribe and payment derived from subscription. Our analysis proceeds in several steps.

We start by building a novel dataset of YouTubers on Patreon along with detailed information on their Patreon revenue and YouTube video output. Through textual analysis of video subtitles and descriptions, we devise methods to characterize the quantity of advertising embedded in videos across different categories.

By leveraging exogenous variation arising from the unexpected data breach, we investigate how a loss of Patreon revenue affects the extent of monetization across different types of advertising. Specifically, we explore how YouTubers adjusted the quantity of advertising in their videos in response to the loss of subscription revenue along four particular margins: 1) in-feed ad breaks; 2) affiliate marketing programs; 3) sponsored content; 4) product placements or reviews.

In general, YouTubers exercise some level of control over these different methods of advertisement. For instance, YouTubers can manipulate the amount of platform-inserted ad-breaks in their videos by utilizing the "10-minute trick" — ensuring their videos are longer than 10 minutes in length (Kerkhof, 2020).[3] Affiliate sales occur when YouTubers earn commission on sales that take place when viewers click through the affiliate links embedded in the video descriptions. Sponsored content refers to videos that are produced in conjunction with an advertiser for the purpose of product or service promotion. Sponsored content may also include videos such as product reviews where YouTubers are provided incentives to cover a product.

To diagnose changes across the broad spectrum of marketing activity, we estimate a difference-in-differences model that compares the differences in the changes to video characteristics after the hack between highly affected creators and those who were relatively unaffected. Our identification strategy assumes that absent the cyber attack, the advertising choices of YouTubers who experienced varying degrees of subscription revenue losses would have evolved along parallel trends. We provide several pieces of evidence supporting this common trends assumption, including visual tests of the pre-trend for all outcomes.

In the aftermath of the Patreon hack, we document a significant shift towards advertising by YouTubers on Patreon that is increasing in the amount of subscription revenue lost. While this general pattern of substitution between subscription and advertisement revenue is perhaps unsurprising, we also document that YouTubers are quite selective in how they include additional adver-

[3] Videos with duration over 10 minutes will contain both pre-roll and mid-roll ads, whereas shorter videos will only contain pre-roll ads.

tising content into their videos. And the extent of the substitution depends critically on the form that advertising took.

Namely, we show that, following the cyber attack, Youtube creators who experienced a larger decline in subscription revenue disproportionately increased the number of affiliate links embedded in their video output. This change arises only after the hack and the timing coincides with that of the subscriber loss.

In contrast, affected YouTubers did not include additional in-feed ad-breaks within their videos through greater utilization of the "10-minute" trick. Specifically, we found no evidence that YouTubers experiencing more substantial subscriber losses lengthened the duration of their videos to above the 10-minute mark so that mid-roll ad-breaks could be inserted. The absence of an effect along this margin shows that the increase in advertising was not uniform across all advertisement formats.

Within the broad category of sponsored content, the effects are also nuanced. On one hand, we detect no discernible increase in the share of videos containing explicitly disclosed sponsored content. We identify sponsorship disclosures through textual analysis of manually generated subtitle captions available for a subset of our videos. Specifically, we search these caption files for disclosure phrases that conform to language recommended by the Federal Trade Commission in their endorsement guidelines for social media influencers. The likelihood that a video contained sponsorship disclosures for YouTubers experiencing larger subscription losses did not increase disproportionately relative to that of YouTubers relatively unaffected by the hack. This suggests that YouTubers did not compensate for the loss of subscription revenue by producing more explicitly disclosed sponsored content.

However, at the same time, we also show that YouTubers who experienced a relatively steeper decline in subscription revenue were more likely to produce videos that featured product reviews. Furthermore, following the hack, their videos also include more references to consumer brands and products which engage in social media marketing campaigns. Therefore, the loss of subscription revenue can facilitate an increase in commercialized content even in the absence of direct sponsorship arrangements.

The main threat to identification in our research design is the possibility that variation in Patreon losses across YouTubers may be correlated with unobserved characteristics of YouTubers that affect the quantity of advertising they choose to embed in their videos. While we attempt to mitigate this concern by including YouTuber fixed effects to account for time-invariant unobservables and also by visually verifying the lack of pre-trends for all the outcomes, this remains a serious issue. To further rule out competing explanations, we implement a secondary research design that compares the evolution of video produced by YouTubers who were on Patreon at the time of the hack with that of late-adopting YouTubers who would join Patreon a year after the hack. Because this strategy uses only variation in the timing of the hack and whether YouTubers were already on Patreon at the time of the hack, it excludes the possible endogenous variation in the number of subscribers lost. Reassuringly, our results using this specification largely accord with those from our main research design.

Overall, our results suggest that a decrease (increase) in subscription revenue leads media providers to increase (decrease) the extent of ad-financing and commercialization of content. However, we find that creators are strategic with respect to the type of advertising they decide to include. We show that YouTube creators increased advertising along the margins which, we argue, are potentially more covert and possibly deceptive in nature: affiliate programs product reviews or placements. Creators did not adjust advertising in ways that were more easily recognized by their viewers. So the extent of substitution between subscription and advertising revenue depends on the format of advertising content and how advertising is incorporated into the content.

Broadly speaking, this paper examines factors influencing media content. We contribute to the empirical literature on endogenous product characteristics in the media industry. Scholars have examined how the variety or quality of media content is affected by supply-side factors such as entry (George and Waldfogel, 2006; Seamans and Zhu, 2014) or consolidations in the media industry (Berry and Waldfogel, 2001; George, 2007), and demand-side factors such as the mix of consumer types (George and Waldfogel, 2003). Furthermore, this strand of literature has explored how media content responds to the emergence of new technology and structural changes in the media industry (Angelucci and Cagé, 2019).

Our research extends the existing literature on several dimensions. First, the comparative statics in this literature typically focus on the effect of advertising revenue on media (Sun and Zhu, 2013; Kerkhof, 2020; Johnson et al., 2023). We study the adoption of a new subscription-based monetization technology on a relatively the media platform YouTube. Our empirical setting emphasizes the significance of the subscriber technology. To our knowledge, our paper is the first to examine the relationship between subscription revenue per se and media content. Our results contribute to recent discussions about the effect of digitization on the media industry. Goldfarb and Tucker (2019) point out that digital technology has disrupted the cost structure of traditional media markets.

Second, there is a long theoretical literature in media economics, dating to Steiner (1952) and Beebe (1977), which focuses on the relationship between revenue structure and program choice. In particular, program selection under advertising-funding is examined in Anderson et al. (2012), Spence and Owen (1977), and Anderson and Gabszewicz (2006), among others. These studies emphasize the inefficiencies of advertising-funded broadcasting. Some recent papers endogenize the advertising-subscription trade-off, with viewers incurring a nuisance cost from advertising (Armstrong and Weeds, 2007; Calvano and Polo, 2020). Stennek (2014) looks at the effect of content exclusivity on the quality of programming.

Finally, many of the theoretical predictions in the media literature hinge on the disutility of advertising to the viewers. Recent advances in advertising technology have reignited a longstanding debate about the distinction between advertising and content in media markets, and how it affects consumers (Sahni and Nair, 2019). Native advertising blurs the line between advertising and consumer-generated content, making it difficult for consumers to distinguish between them. Anecdotal evidence supports the notion that consumers often mistake native advertisements for independently created editorial content (Bakshi, 2015). While there exists a literature on advertising content, with notable exceptions, research on native-advertising remains scarce (Anderson and Renault, 2006, 2013; Chatterjee and Zhou, 2017).

The masked nature of advertising via user-generated content raises important questions for policymakers and researchers alike. For instance, Pfeuffer et al. (2021); Johnson et al. (2021) study influencer marketing and effect of sponsorship disclosure on YouTube. This paper contributes to this discussion by examining the determinants of sponsored content on a significant digital media platform. By examining the effect of subscription shocks on more and less observable formats of advertising, we show how the emergence of content marketing provides media producers additional levers in optimizing the trade-off between revenue generated by subscribers and advertisers.

The rest of the paper proceeds as follows. In Section 2, we provide background information on YouTube, advertising mechanisms, and Patreon. In Section 3, we describe our data sources. In Section 4, we provide descriptive statistics. In Section 5, we define our empirical strategy and in Section 6, we present our main results. We probe the robustness of our findings in Section 7 Finally, in Section 8, we conclude.

2 Background

2.1 YouTube

YouTube is the world's largest video platform. YouTube was founded in February 2005, and was acquired by Google in November 2006. As of February 2017, over one billion hours are watched on YouTube on a daily basis. As a point of reference, according to Nielsen, American viewers watch about 1.25 billion hours of television per day.[4]

While recently, YouTube has begun producing and distributing original content, historically, almost all content creation on YouTube was user-generated. An individual who uploads videos to YouTube is typically called a "YouTuber". YouTube pays all web server costs associated with storing and streaming videos. In the past, YouTube had some limitations on the maximum size of video that could be uploaded, but currently, there are no such restrictions. YouTube has some restrictions on the types of content which can be uploaded.[5]

[4] YouTube Tops 1 Billion Hours of Video a Day, on Pace to Eclipse TV (*Wall Street Journal*, 2017-02-27)

[5] For example, pornography, gore, or copyright-infringing content are not permitted.

Initially, YouTube videos did not have advertisements. However, beginning in 2007, and increasingly through the rest of the 2000's and 2010's, YouTube began to show advertisements before and during YouTube videos. YouTube instituted a partnership program with its content creators to incentivize continued content creation. Partnered video YouTubers received advertisement revenue on the basis of Cost per Thousand Impressions (CPM).

Currently, YouTubers receive around 55% of advertising revenue generated by their videos, while YouTube keeps the remaining 45% of advertising revenue.[6] YouTubers have some discretion on the degree of monetization of their videos, but they do not select the exact advertisements shown on the videos. Which ads are inserted is determined algorithmically and YouTubers have little direct control over that process.[7]

These in-feed ad breaks can play before the video (pre-roll) and also midway through the video itself (mid-roll). Videos with duration over 10 minutes will contain both pre-roll and mid-roll ads, whereas shorter videos will only contain pre-roll ads. As a result, one way for YouTubers to manipulate the number of YouTube ads shown is to utilize the "10-minute trick" — ensuring their videos are longer than 10 minutes in length (Kerkhof, 2020).

2.2 Content Marketing

In addition to advertising inserted by the YouTube platform, YouTubers also have the discretion to include additional advertising within their content.

As a social media and video-sharing platform, advertisements and consumer-generated content on YouTube are often intermixed (Campbell and Grimm, 2019). Two primary forms of content-marketing taking place on YouTube are affiliate marketing and sponsored content.

2.2.1 Affiliate Marketing

Affiliate programs allow content creators to receive a commission for products purchased by consumers who arrive at the platform via the creator's content. Most large ecommerce platforms

[6] See: https://finance.yahoo.com/news/much-youtubers-213800157.html

[7] The particular advertisements shown and the CPM are determined through an auction mechanism on the Google Ads platform, where advertisers bid on ads targeting particular user or video characteristics.

have affiliate programs, including Amazon, eBay, Walmart, and Best Buy. Smaller businesses also often have affiliate programs. WooCommerce and Shopify, popular small business ecommerce software, both come with affiliate marketing extensions allowing small businesses to implement their own affiliate programs.

One of the largest affiliate program is the Amazon Associates Program.[8] Affiliates are given a unique "affiliate link" to use when linking to Amazon. When a consumer purchases a product via the affiliate link, affiliates receive between 1-10% of a product's sale price, depending on the product's category.[9] Currently, the program has over 900,000 affiliates.

YouTubers can include links both in the summaries of their videos, and within their videos as text boxes that appear within the video. While most affiliate programs require some disclosure of an affiliate relationship when an affiliate provides their affiliate link, in practice, affiliate relationships are rarely disclosed by YouTubers (Mathur, Narayanan and Chetty, 2018).

2.2.2 Direct Sponsorship

In addition to affiliate marketing, YouTubers may receive direct payments in exchange for producing content featuring a brand or product as part of a promotional campaign.

For instance, YouTubers can promote a sponsor's product or services during a dedicated segment within their video. Alternatively, a YouTuber may get paid to prominently feature a particular product within a video.[10] A less conspicuous form of direct sponsorship is product reviews - where YouTubers either receive payment for reviewing a product, or receive a product for free in exchange for a review.

Because, sponsored content can resemble regular user-generated content, these types of advertisements are arguably less obtrusive and possibly more covert to viewers than ads through the YouTube platform. To ensure more clear delineation between ads and content and to prevent deceptive marketing, the FTC provides "Endorsement Guidelines" governing how sponsorship needs

[8] https://sellercentral.amazon.com/gp/help/external/G28591?language=en_US

[9] When someone visits a platform through an affiliate link, the platform's webserver will return a "cookie" which gets stored in the visitor's web browser. The affiliate gets credited with a purchase made before the cookie expires, or if the visitor arrives at the viewer's website via some other affiliate's link. Amazon's cookie expires after 24 hours.

[10] This video from the cooking YouTuber "Binging with Babish" is an example of this form of sponsored content.

to be disclosed, and in recent years, the FTC have litigated several firms engaging in deceptive advertising.[11] However, during the time period we study (2015-2016), the issue of undisclosed sponsorship and branded content on YouTube was frequently discussed in the media (see footnote for three examples).[12]

This possibly prompted the FTC to publish clarification regarding influencer marketing in 2015. However, a recent law review article concludes that, in practice, social media influencers are rarely held directly liable for misleading advertisements (Bentz and Veltri, 2020).

Sponsorship relationships may be negotiated simply through individual bilateral communication between a content creator and a company, but these relationships can also be established and negotiated through dedicated platforms to facilitate sponsorship relationships.[13]

2.3 Patreon

Until relatively recently, advertising was the main way through which YouTube creators were able to monetize their content. The launch of Patreon in 2013 provided an alternative source of revenue for YouTubers. Patreon is a membership platform that provides business tools for creators to run a subscription content service. Patreon streamlined the process of collecting payments on a recurring-basis from viewers for YouTubers without the need for any technical/web development expertise. Prior to the founding of Patreon, content creators could potentially create their own ad-hoc subscription monetization solutions. For instance, by including a Paypal link in their video descriptions. However, these ad-hoc methods can be more difficult to set up and maintain, imposing greater costs on YouTubers. They also do not have as much visibility and ease of use as compared to the dedicated Patreon pages.

On Patreon, online content creators create profile pages on which they can solicit subscription payments from viewers. Creators can create a tiered reward structure. For instance, there are

[11] FTC Endorsement Guides

[12] https://www.digitaltrends.com/gaming/ftc-orders-warner-bros-to-disclose-sponsored-youtube-content/; https://www.eurogamer.net/blurred-lines-are-youtubers-breaking-the-law https://www.osborneclarke.com/insights/undisclosed-advertising-on-youtube-a-potential-pitfall-for-games-brands

[13] One of the oldest platforms, Traackr, was created in 2006. The platform tracks online content creators, their social media posts, and demographics of their followers, allowing brands to target influencers precisely. Other platforms also facilitate the negotiation of contracts with influencers.

membership tiers that provide token rewards, such as themed desktop wallpapers, and the ability to comment on livestreams. Some other common rewards given by other creators are mentions of patrons' names at the end or in summaries of videos, early access to videos, or access to "bloopers" during video production. The profile page shows the creator's *current* monthly revenue and total number of patrons. While all profile pages show the total number of patrons, since February 2017, creators have been able to hide their monthly revenue. All data in our sample is from before February 2017, and so we will have revenue data in all months for all creators in our sample.

Patreon is used by YouTubers, webcomic artists, writers, podcasters, musicians, and other categories of creators who regularly publish their content online. Figure A1 shows the number of Patreon creators by category. The largest single category of creators are video creators, most of whom are YouTubers. Other large categories are bloggers, podcast hosts, and adult content creators. Additionally, there are not significant changes in the category composition of Patreon creators over time.

Patreon charges a "platform fee" of between 5 and 12% of subscription payments taken by creators, and a "payment processing" fee of between 3 and 5%.[14] As of 2019, Patreon has over 100,000 active creators, 4 million monthly patrons, and paid out approximately $500 million dollars annually to creators[15]. In recent years, Patreon has expanded into offering business software suites for creators to manage transactions.[16]

2.3.1 Patreon Hack

In October 2015, Patreon was the target of a large-scale cyber-attack. The website database, with nearly 15 gigabytes of password data, donation records, and source code was compromised and exposed. More than 2.3 million email addresses and private messages between patrons and creators were also leaked in this breach. The hackers subsequently posted the user information online in a data dump.

The cyber-attack was widely reported at the time. However, the nature of the data breach was

[14] https://support.patreon.com/hc/en-us/articles/360027674431-My-earnings-fees
[15] Crunchbase
[16] See: https://techcrunch.com/2019/02/12/patreon-product/

initially unknown and misinformation regarding the extent of the hack spread on social media.

Following the data breach, it was reported many patrons received extortion emails for money in exchange for the protection of personal information.[17] In actuality, no data posing a significant privacy risk was made publicly available. While passwords and privately identifiable information such as credit card information, home addresses, and Social Security numbers were in the leaked database, these data were stored encrypted.[18] However, some patrons may have been unaware of exactly what data had been leaked due to misinformation on social media. These extortionists scams became widely reported in the media by November 2015.

3 Data

We make use of two main data sources. First, we use information on the number of subscribers and amount of subscription payments from a Patreon tracking site. Second, we use video metadata and subtitles scraped from YouTube. Notably, we do not make any use of the data acquired in the Patreon hack - we only use the hack indirectly as a shock.

3.1 Patreon Revenue

We acquire information about Patreon from *Graphtreon*, a website tracking Patreon creators. A Patreon creator's profile provides the creator's *current* monthly revenue and/or the creator's *current* number of subscribers. However, a Patreon creator's profile does not contain information about *past* revenue or *past* number of subscribers.

Therefore, Graphtreon scrapes each Patreon creator's monthly revenue and number of subscribers on a daily basis in order to construct a time series of each creator's history of daily revenue and number of subscribers. We collect the revenue and number of subscribers of every Patreon creator on Graphtreon.

Graphtreon first began scraping revenue and subscribers information on March 18, 2015 (nearly 2 years after Patreon's start in 2013). Initially, Graphtreon did not scrape the universe of Patreon

[17] See: https://www.businessinsider.com/patreon-users-targeted-by-extortion-scam-crowdfunding-2015-11
[18] Names, email addresses, messages, and subscription information were stored in plain text.

creators, but rather a small subset. Gradually, Graphtreon expanded the set of creators for which it scraped data.[19] By March 2016, Graphtreon had full coverage of Patreon.

Our sample consists of YouTubers who joined Patreon before May 2015. In order for us to obtain information on Patreon revenue from Graphtreon, we are additionally constrained to creators who began being scraped by Graphtreon before June 2015. As a result, our sample consists of YouTubers who were on Patreon as of May 2015 and whose information was available on Graphtreon as of June 2015.

This is undoubtedly a selected sample of YouTubers. However, since our empirical strategy relies only on comparisons of YouTubers within our selected sample, this type of selection bias should not affect the internal validity of our results. For our sample, we observe each YouTuber's monthly Patreon revenue on a daily basis from the period of June 2015 onward.

3.2 Video Data

We use the YouTube Data API to scrape metadata about videos created by YouTubers in our sample. Specifically, for videos created between August 2015 and July 2016, we retrieved a video's metadata that includes the number of likes, dislikes, views, duration, and category of each video. Additionally, we use the summaries and captions of videos to create two main outcomes of interest: the number of affiliate links, and the occurrence of sponsored content.

3.2.1 Affiliate Links

YouTube videos' descriptions often contain outgoing links. Some of these links are affiliate links, links through which the link provider can collect a commission. Each major eCommerce provider's affiliate links have a regular structure that allows identification of the affiliate link. For example, a Walmart affiliate link will contain the string "selectedSellerId".

We parse all links in the summaries of videos and identify affiliate links from several large retailers that use referral programs, including Walmart, Amazon, Autodesk, etc. We also parse some additional affiliate links if the link contains the string "affiliate_id", which is often used in

[19] From Figure A2, we see that there are several spikes in the number of creators on Graphtreon. These spikes reflect updates of Graphtreon's web scraping procedures to capture a greater proportion of the creators on Patreon.

affiliate URLs of smaller eCommerce providers.

3.2.2 Captions & Video Titles

We extract further information on the content of published videos through text-based analysis of the closed captions. For any YouTube video, the video creator can upload captions that transcribe the speech or dialogue in the video to text.[20]

In total, there are 2,475 videos in our sample with manually-generated captions. This is a small subset of the overall sample of videos. To mitigate concerns of selection bias associated with using a sub-sample, we restrict our attention only to YouTuber creators who provided subtitles for at least one video before and after the cyber-attack. By excluding YouTubers who started or stopped providing subtitle files after the cyber-attack, we hope to alleviate the possibility that differential changes in caption availability can drive the result. We further address this issue in the robustness section.

We measure sponsored content using these caption files. We identify within captions occurrences of particular phrases that YouTubers often use to disclose sponsorship relationships. Common examples of these disclosure statements include "the sponsor of this video", "this video is sponsored by", "thank our sponsors", etc. followed by the identity of the sponsor. The basket of disclosure phrases can be found in Appendix A. Using these dictionaries of disclosure phrases, we create measures for explicitly disclosed sponsored videos based on whether the caption files contain such a phrase.

A limitation of this procedure is that sponsorship relationships may exist even without explicit disclosures. To gauge the extent of this type of behavior and quantify the overall commercialization of video content, we construct two additional variables.

First, we search for mentions of a large set of consumer brand names and product names that may sponsor YouTubers because such references may be indicative of an underlying financial arrangement. Our list of brands comes from the Social Index on the website BrandWatch, which is

[20] YouTubers can also enable automatically generated captions using Google's Text-to-Speech API. However, we found that the accuracy of the automatically generated captions was poor and could not be reliably used in the text analysis. Therefore, we focus on videos with manually uploaded caption files.

a market research company that tracks the social marketing presence of consumer product companies.

Second, within the set of videos , we identify whether videos feature product reviews by observing whether the title of a video contained the word "review". We construct a variable that equals one if a video include review in its title.

3.2.3 Youtube Ad-Breaks

Unfortunately, we do not observe the precise number of ad-breaks in Youtube videos nor the amount of advertising income YouTubers receive from those ad-breaks. However, YouTube's monetization policy has a distinctive and notable feature: a discontinuity in the mapping from a video's duration to the number of ad breaks that YouTube permits. Namely, if a video is shorter than ten minutes, YouTube allows exactly one ad break (pre-roll) to be run at the beginning of the video. Videos longer than 10 minutes can have additional *mid-roll* ads which are shown in the middle of the video.

Given this, we use the video duration as another measure of advertising revenue directly from YouTube. The significance of this 10-minute threshold is discussed extensively in Kerkhof (2020). Kerkhof (2020) utilizes the introduction of a new ad-break tool in November 2015 as an informational shock for YouTubers regarding the existence of this 10-minute cutoff. But as Kerkhof (2020) points out, the actual discontinuity existed before that date. Given that we are studying a small sample of professional YouTubers who earn significant amounts of income through their YouTube endeavors, it seems unlikely that they would be unaware of the 10-minute trick until the release of the new ad-break tool. As we will show in the next section, for the videos in our sample there is significant bunching in video duration right at the 10-minute mark, indicating that the YouTubers we study were familiar with this discontinuity.

To get a sense of the magnitude of advertising revenue on the YouTube platform, we provide some suggestive evidence on Patreon video creators' YouTube advertising revenue using data from *Social Blade*, a website that estimates YouTubers' earnings from YouTube.

Social Blade constructs its estimates using detailed information on the video views over time along with reasonable approximations of YouTube's CPM rates (Clicks per thousand views). From surveys of YouTubers, Social Blade found that advertising earnings on YouTube tend to vary between \$0.25 CPM and \$4.00 CPM, and so it presents bounds on advertising earnings using those CPM values.

4 Descriptive Statistics

First, to understand the effect of the Patreon hack, we plot the total and per-creator weekly earnings over time. Figure 1 shows a dramatic reduction in both earnings metrics.

Next, in Table 1, we provide descriptive statistics on Patreon earnings for our sample of YouTubers who joined Patreon prior to May 2015. We refer to this group as "early adopters". This is the sample of YouTubers we study in the majority of our analysis.

Among the early adopters, the median Patreon video creator makes \$780/month, whereas average earnings are \$2336/month, which reflects the fact that earnings are skewed to the right tail due to the existence of superstars and outliers earning in excess of \$10,000/month. Additionally, the creators in our sample have had Patreon profiles on average for one year.

We also provide some limited evidence on early Patreon adopters' YouTube advertising earnings using Social Blade estimates. Social Blade multiplies channels' views in the last month by a lower bound and upper bound of potential CPM to bound advertising earnings. We use the midpoint of these bounds to create earnings estimates.

As shown in Table 1, for these YouTubers, there appears to be an approximately even split of revenue coming from subscriptions and from advertising. However, this likely understates the total advertising revenue received by YouTubers, as these figures exclude direct payments from advertisers for sponsored content or affiliate marketing related revenue.

To quantify the degree to which YouTubers monetized their videos through YouTube, we examine their utilization of the 10 minute "trick". As shown in Figure 2, the distribution of duration of videos produced by YouTubers in the sample exhibits significant bunching just above the 10-minute mark. Since the YouTubers have near-perfect control over their videos' duration, this suggests YouTubers were aware of the 10-minute policy and strategically manipulated the length of their videos in order to maximize advertisement revenue from YouTube.

4.1 Selection into Patreon

Which media producers adopt subscription technology? In order to better understand selection into Patreon, we compare the YouTube videos produced between August 2015 and July 2016 by YouTubers who were already on Patreon by June, 2015 (early adopters) to a complementary sample of videos from August 1, 2015 and July 31, 2016 of Patreon "late adopters" who joined the platform between August 1, 2016 and January 1, 2017. In other words, we compare the video content of YouTubers who already joined Patreon at the time of the video's production and of those who had not.

Comparing early and late adopters in Table 1, we find evidence consistent with the notion that there was positive selection into Patreon: higher quality YouTubers were early to adopt this new subscription technology. From Table 1, we see that YouTube videos in both samples tend to have many more likes than dislikes, but that the Patreon creators in our early adopter sample have a better like-to-dislike ratio and significantly higher number of likes overall. The videos produced by the early adopter sample also tend to have substantially higher viewership. Furthermore, videos produced by early adopters have a much greater level of audience engagement, as measured by the number of comments. These differences indicate that early adopters possibly produced higher-quality videos during the same time period.

In terms of video duration, both samples' videos have comparable lengths. The average duration for late adopters is approximately 20 minutes with a median length of 10 minutes while the average duration of early adopters' videos is slightly longer at 23 minutes.

5 Empirical Strategy

Our goal is to identify the effect of subscription revenue on the content-creating behavior of Patreon content creators. Our main identification challenge is that the amount of subscription revenue received by YouTubers is endogenous. In order to produce credible estimates, we need exogenous shifts in YouTubers' Patreon earnings that are unrelated to viewers demand or preferences for video content. The 2015 Patreon cyber-attack provides such a source of identifying variation.

In the data, we observe that many subscribers canceled their recurring Patreon payments following the hack and the exposure of user information. While canceling a Patreon subscription had no effect ex-post on whether or not a subscriber's information was leaked. This decision could reflect either the incorrect belief that canceling their subscription would protect their private information, or an acquired distaste or distrust for Patreon due to the security breach.

From Figure 1, total monthly payment for our sample of YouTubers on the Patreon platform, in the aggregate, declined by approximately $400,000 in the months following the of cyber-attack. On average, each Patreon creator experienced a revenue loss of approximately $120 per month after the cyber attack was publicized.

However, Figure 1 does not speak to whether there is a variation in the amount of revenue lost across Patreon creators. To assess this, we estimate a first-difference model of the following form:

$$Earnings_{it} = \beta Post_t + \delta_i + \varepsilon_{it} \tag{1}$$

where $Earnings_{it}$ are the weekly Patreon earnings of YouTuber i at time t measured in levels and logs. $Post_t$ is a dummy variable indicating the post-cyber-attack period.

To illustrate the results graphically, we plot the evolution of Patreon earnings over time in Figure 4. We separate the sample of YouTubers into those who were most adversely affected by the hack (upper quartile in losses) and those who were not (bottom quartile in losses).

These figures are shown in Figure 4. In the first figure, we observe that on average YouTubers experienced a 20-30% decline in subscription revenue following the Patreon hack. However, the decrease in Patreon subscription revenue varied significantly across content creators. When dividing the losses into quartiles, we see that the upper quartile of losses represented close to 50% of revenue whereas YouTubers in the lower quartile of losses were nearly unaffected. The main driver of this variation is likely the degree of privacy concerns of different creators' subscribers.

We exploit variation in the size of the shock to study the causal relationship between subscription revenue and content choices in a difference-in-differences framework. Specifically, we compare the advertising shift before and after the cyber-attack, produced by YouTubers who experienced large revenue losses with that of YouTubers whose Patreon revenue was relatively less affected by the cyber-attack.

The identification strategy relies on two key assumptions. First, the Patreon hack does not affect content creators' behavior except through the change induced in Patreon subscription revenue. The second assumption is the parallel-trends assumption: absent the hack, the video content of YouTubers who experienced different amounts of revenue losses would have evolved along similar trends.

We think these assumptions are plausible given that the hack was platform-wide and not individual YouTuber-specific. A subscriber's willingness to cancel his or her Patreon account is going to be largely dictated by the concern for data security on the Patreon platform as opposed to preferences for any single YouTuber. Ex-ante there is little reason to expect the concern for data security would be correlated with video characteristics of the particular creator with whom they are subscribed.

Furthermore, even if the degree of privacy concerns of different creators' subscribers was correlated with the video content of YouTubers to whom they are subscribed, this correlation would not invalidate our research design as long as it remained time-invariant and not evolving over time in a way that is independent of the cyber-attack.

In our empirical analysis, we take steps to address this concern and assess the plausibility of our research design. In particular, we demonstrate constant pre-trends in several content-related outcomes between creators experiencing large revenue losses (i.e. those with highly privacy-conscious subscribers) and creators experiencing low revenue losses (i.e. those with less privacy-conscious subscribers).

5.1 Specification

We formalize the empirical strategy outlined above here. As a first step in the empirical analysis, we define a continuous measure of *treatment* based on the amount of monthly Patreon revenue a YouTuber lost as a result of the hack (i.e. the decline in monthly Patreon revenue from December 2015 to February 2016): $Loss_i$. We then relate changes in video characteristics to the magnitude of the loss. We pursue a generalized difference-in-differences strategy. Specifically, we estimate the following baseline regression model:

$$y_{vit} = \beta Loss_i * Post_t + \sigma_i + \sigma_t + \varepsilon_{vit} \tag{2}$$

where v denotes a YouTube video; i and t index the creator and the month-year that the video was published, respectively. $Post_t$ is an indicator variable equaling one if the video was produced after December 2015. This choice of timing is motivated by the earnings time-series shown in the previous section where we observe that the earnings decline on Patreon materializes during December 2015.

$Loss_i$ is the amount of Patreon revenue lost by YouTuber i between December 2015 and February 2016 measured in thousands of dollars. y_{vit} is the characteristics of video v created by YouTuber i and published month-year t. For the majority of our analysis, the dependent variable is a binary indicator equal to one if the video contains advertising content.

Thus, we estimate Linear Probability Models where we compare the relative change (before and after the hack) in the likelihood of including advertising in videos between creators who experienced a greater loss in subscription revenue and those who experienced less. The parameter of interest, β, captures the percentage change in the probability of including advertising attributed to a percentage increase in subscription revenue lost.

6 Main Results

In this section, we investigate the causal effect of subscription revenue on advertising strategies pursued by YouTubers.

6.1 Effect of Hack on Patreon Earnings

To start, we revisit the first difference model of Equation 1, which shows the evolution of Patreon earnings around the time that the cyber attack on the platform took place and became widely publicized. Table 2 presents the results of this "first-stage" regression. Each observation is a creator and month-year pair. Columns 1 and 2 show the effect of the Patreon hack on income from Patreon. The effect size is considerable. The estimates suggest that the hack was associated with an over 20 percentage points reduction in average monthly income. The results in levels are displayed in columns 3 and 4. Accordingly, we find that income declined by over $140 following the hack after accounting for YouTuber fixed effects. This is a sizable loss, corresponding to approximately 18% of the median monthly Patreon earnings for YouTubers in our sample.

To provide some additional context for the impact of the hack on subscriber loss, we also devise a secondary test of this first-stage relationship. We observe that YouTubers often appeal to their viewers for Patreon support. Specifically, at the beginning or end of their videos, YouTubers will, at times, directly ask viewers to join Patreon in order to support their channels. YouTubers will also thank their existing Patreons. Naturally, these types of solicitation might be affected by the loss of Patreon revenue. To measure the frequency of this behavior, we search for mentions of "Patreon" in the caption files and create a binary variable equalling one if the caption files contained a reference to Patreon.

In Panel A of Figure 3, we illustrate the evolution in the share of YouTube videos that mention Patreon over time. Specifically, we regress the dummy for whether a video referenced Patreon on month-year dummies for the subset of videos with manually generated caption files. We plot the time series and from the figure, we observe that Patreon references appear to increase dramatically following the cyber attack and the timing coincides with that of the loss in subscription revenue.

Furthermore, we show that the increase in Patreon mentions was concentrated amongst YouTubers who experienced the largest decline in subscription revenue. Specifically, we regress the outcome on the interaction of the month-year dummies and the loss in subscription revenue. We plot these coefficients in Panel B of Figure 3. Following the cyber attack, the coefficients on the interaction terms spike up. This indicates that increases in solicitations for additional Patrons were increasing in the amount of revenue lost. This is consistent with YouTubers attempting to persuade their viewers to join Patron in order to offset the lost subscription revenue.

These results both serve as a sanity check for the accuracy of the manual subtitle files and also illustrate that YouTubers were aware of the loss of Patreon revenue. The salience of the subscription loss provides much-needed context for the strategic responses we document in the next section.

6.2 Effect of Earnings Loss on Advertising

To motivate our main regression specifications, we present event study figures for the effect of lost Patreon revenue on four categories of advertising content: 1) affiliate marketing (measured in both the number of affiliate links as well as the share of all URLs containing affiliate links); 2) the extent of ad-break monetization as indicated by whether the video duration exceeds 10 minutes; 3) sponsorship disclosures; and 4) the overall level of commercialization as evidenced as whether a video features product reviews and the number of references to consumer products and brands it contains.

These figures are shown in Figure 5. In the first two figures, we observe that the loss of Patreon revenue was associated with an increase in the number of affiliate links as well as the share of all URLs embedded in the description that contained affiliate links. The coefficients suggest that YouTubers who lost more subscription revenue due to the hack, in turn, included more affiliate links and product mentions in their YouTube videos. The timing of the increase in these forms of advertising coincided with that of the Patreon hack.

For sponsorship disclosures and whether video length exceeds ten minutes, we do not observe any visible trend break coincident with the cyber-attack. Finally, in the last two figures, we observe there is a rise in the overall commercial content of videos as evidenced by the increased references to consumer brands and products in the caption files as well as whether videos featured product reviews.

Reassuringly, we observe little evidence of pre-trends across all outcomes. For each manner of advertising, the difference in the level of advertising between relatively more affected and less affected YouTubers was stable prior to the hack. This lends credibility to the common-trends assumption and provides evidence in support of a causal relationship.

To explore this pattern more systematically and better quantify magnitudes, we estimate Equation 3 to show the effect of subscription revenue loss on quantity of advertising by type. We start with the number of affiliate links embedded in the video description. These main findings are summarized in columns (1)-(4) of Table 3. Columns 1-2 show the results on the number of affiliate links across two distinct specifications with different sets of fixed effects.[21] In columns 3-4, rather than using the number of affiliate links as the outcome, we consider the percentage of all URLs in the video description containing affiliate links.

The analysis is organized at the video level. The positive coefficient on the $Post \times Earnings\ Loss$ interaction variable indicates that creators who experienced greater loss of subscription revenue, on average, added more affiliate links to their videos after the hack. This is consistent with the notion that these creators responded to the loss of subscription revenue by substituting toward more revenue via affiliate marketing.

Interestingly, columns 5 and 6 show there is a statistically significant negative effect on the likelihood that the video is over ten minutes long. As detailed in Section 5, ten minutes is the cutoff for multiple ad-breaks to be inserted in the video on the YouTube platform. From Figure 2, we saw that there is significant bunching at the ten-minute duration because of strategic manipulation around this discontinuity. Surprisingly, we do not see increased bunching around the ten-minute mark in a way that corresponds with subscription revenue loss. This indicates that YouTubers prefer

[21]We winsorize the number of affiliate links at 15 (99.5 percentile) throughout our analysis to mitigate concerns of outliers.

to monetize by way of affiliate programs as opposed to YouTube inserted ad-breaks. This could be due to the fact that YouTube-inserted ads are more observable to viewers and perhaps impose greater nuisance costs.

Table 4 shows the effect of the Patreon hack along three additional margins, each constructed using videos' manually generated captions. Columns 1 and 2 demonstrate that the share of videos with sponsorship arrangement (as indicated by an explicit disclosure) did not change after the hack in a manner that was proportional to the subscription revenue loss experienced by the YouTubers.

In columns 3 and 4, we show that mentions of consumer brands and products increased in videos produced by YouTubers who experienced large Patreon losses following the hack. This suggests that YouTubers substituted towards advertising and increased the overall commercialization of content as a result of the loss in subscription revenue. Along similar lines, columns 5 and 6 show that YouTubers who experienced larger Patreon earnings losses were also disproportionately more likely to produce product reviews.

7 Additional Results & Robustness

7.1 Extensive Margin

We examine whether our results could be explained by effects on the extensive margin, namely whether YouTubers adjusted the number of videos produced in response to losses in subscription revenue. In order to characterize the effect on the number of videos, we organize our analysis at the creator by month-year level. This allows us to examine the impact of the hack on the monthly number of YouTube videos. The results are presented in Table 6. We measure the number of videos in both logs and levels and across both specifications, we find very little evidence that the number of videos changed after the cyber-attack for YouTubers exposed to subscription loss. In other words, YouTubers did not decrease or increase their video output in response to the loss of subscription revenue.

7.2 Selection Bias in Subtitle Availability

Several of our outcome variables are constructed using caption files uploaded by YouTubers, one concern is that changes in subtitle availability may drive our findings. In particular, if the YouTubers become more (or less) likely to provide subtitle files after the hack, that could lead to mechanical correlations which may confound our findings.

In our main specifications, we attempt to address this by restricting our sample only to YouTubers who included captions for at least one video before and after the hack in order to ensure that our comparison is driven by a relatively consistent sample of YouTubers. However, this approach may not completely solve the problem. Therefore, we also test whether the hack affected the likelihood of providing captions in order to assess the urgency or potential impact of the selection bias.

We create a dummy variable that equals 1 if the video creators had uploaded caption files. We analyze this outcome through the difference-in-differences model using the sample of YouTubers who included caption files in at least one video prior to and after the hack. The results are presented in Table 7. The coefficient on the interaction between *EarningsLoss* and *Post* is statistically insignificant and small in magnitude. This indicates that YouTubers disproportionately exposed to the shock did not become more or less likely to provide captions. This suggests that selection bias with respect to subtitle availability is unlikely to drive our findings.

7.3 Alternative Specification

Our main difference-in-differences model exploits the comparison between YouTubers on Patreon who experienced large subscription revenue losses and those who did not. As such, our treatment — revenue loss arising from the hack — is continuous. As discussed recently in Callaway, Goodman-Bacon and Sant'Anna (2021), the extension of difference-in-differences research design to settings with continuous treatment is far from trivial and the use of two-way fixed models in these situations is problematic and requires stricter assumptions than standard common trends assumption.[22]

To address these concerns, we formulate a secondary research design that sidesteps this issue by employing a distinct control group — Patreon late adopters. As described in Section 4.1, we refer to Patreon late adopters as YouTubers who joined Patreon between August 2016 and January 2017. To be clear, the late adopter group consists of YouTubers who have yet to join Patreon by the time of the hack and should therefore be unaffected by the hack.

By using this group as a control, we are able to estimate a model that closely resembles the canonical difference-in-differences framework where we examine the changes in video content of treated YouTubers (early Patreon adopters) before and after the hack as compared to the changes of videos produced by untreated YouTubers (late Patreon adopters). This approach is summarized by the following regression specification:

$$y_{vit} = \beta \cdot EarlyPatreon_i \times Post_t + \sigma_i + \sigma_t + \varepsilon_{vit} \tag{3}$$

where the sample consists of all YouTube videos v produced by our control and treated group between August 2015 and July 2016.[23] $EarlyPatreon_i$ equals 1 if YouTuber i joined Patreon prior to May 2015 (early adopter), indicating that their videos were produced while they were on Patreon. Conversely, $EarlyPatreon_i$ equals 0 if YouTuber i belongs to the late adopter group and their videos were produced prior to them joining Patreon.

One data limitation we face with this approach is that we do not have subtitle files for the videos produced by late adopters of Patreon. Therefore, our analysis in this section is restricted to outcomes we create using video meta-data: affiliate links and YouTube ad-breaks (video length/duration). However, because of the significantly larger sample size, we are more sufficiently powered to detect small but meaningful effects along those dimensions we can measure.

[22] In our case, the additional assumptions required would be the effect of revenue loss to be equal for people receiving different dosages of treatment (i.e. experiencing different levels of loss) at the same dosage level. See: https://causalinf.substack.com/p/continuous-treatment-did for an extended discussion.

[23] Note that our sample excludes, by construction, YouTubers who joined Patreon between May 2015 and June 2016 because their treatment status would be mixed.

We start by examining the number of affiliate links and the extent of YouTube ad-monetization, as proxied by whether the video length exceeds 10 minutes. We produce event-study figures corresponding to these two outcomes in Figure 6. The first figure shows that, following the cyber-attack, YouTubers on Patreon dramatically increased the number of affiliate links embedded in their videos compared to YouTubers not on Patreon at the time. Given that YouTubers on Patreon lost subscription revenue on average, this is consistent with our earlier finding and suggests that a drop in subscription revenue leads to an increase in affiliate marketing.

The second figure show that there appears to be a small decline in the number of video above 10 minutes following the hack. This is consistent with our earlier results which showed a negative point estimate on the effect of revenue loss on videos longer than 10 minutes.

Results using this specification are presented in Table 6. Overall, our findings using this alternative approach are qualitatively similar to our earlier results. We find subscription revenue loss leads to a statistically significant increase in affiliate marketing but no changes in the extent of ad-break monetization through utilization of the 10-minute trick.

8 Conclusion

Throughout its history, the media industry has undergone a series of social and technological transformations which have significantly impacted media content and the choice thereof. For example, in most of the 19th century, American newspapers followed a subscription-based model, and had explicit party affiliations which constrained the stories editors and journalists would cover. With the growth of advertising markets in the late 19th century, newspapers became less reliant on partisan support and featured less political bias in an effort to attract advertising revenue (Petrova, 2011).

The norm of ad-revenue-funded media became firmly established in the 20th century. Against this backdrop, recent technological innovations in content delivery have enabled digital media creators to gate their content and put them behind a paywall, allowing them to monetize their consumers more directly. This has facilitated a return to a subscription-based model in many areas of digital media. The re-emergence of subscription revenue models represents a dramatic reversal from how the media industry has been organized in recent history.

In this paper, we used an exogenous shock to the subscription revenue of YouTubers on Patreon to study how subscription revenue affects the quantity and type of advertising embedded in their video content. We focus on Patreon video creators because rich metadata from YouTube allows us to study multiple dimensions of content quality.[24]

Our analysis exploits the timing of the 2015 cyber-attack on Patreon which had a significant negative effect on earnings received by YouTubers on the Patreon platform. Using a difference-in-difference approach, we found that a decrease in subscription revenue leads to more affiliate links in the summaries and descriptions of videos. YouTubers also increased the commercial content in their videos as evidenced by references to consumer brands and product reviews.

On the other hand, we note an absence of an increase in advertising through other means, namely explicitly disclosed sponsored content as well as in-feed ad-breaks. If anything, we find that YouTubers may have responded by reducing YouTube ad-breaks, perhaps to offset the increase in advertising along other dimensions. We argue that these contrasting effects could be explained by the fact that affiliate links and product placements are more covert forms of advertising that are less observable to viewers than disclosed sponsored content or in-feed ad-breaks.

A possible conclusion to draw from our analysis is that the extent of substitutability between advertising and subscription depends on the format and method of advertising and changes in subscription revenue may induce changes, not just on the level of advertising, but also on the advertisement mix within media content.

[24]A promising direction for future research is to extend our analysis and see if similar dynamics exist with other media influencers. Additionally, one important aspect of digital media is the role of media aggregators in determining content viewed by consumers. We would like to study how these aggregators interact with economic incentives to determine the type of content YouTube creators produce (Athey, Mobius and Pál, 2017).

References

Anderson, Simon P, and Jean J Gabszewicz. 2006. "The media and advertising: a tale of two-sided markets." *Handbook of the Economics of Art and Culture*, 1: 567–614.

Anderson, Simon P, and Joshua S Gans. 2008. "Tivoed: The effects of ad-avoidance technologies on broadcaster behaviour." *Available at SSRN 1295046*.

Anderson, Simon P, and Régis Renault. 2006. "Advertising content." *American Economic Review*, 96(1): 93–113.

Anderson, Simon P, and Régis Renault. 2013. "The advertising mix for a search good." *Management Science*, 59(1): 69–83.

Anderson, Simon P, Øystein Foros, Hans Jarle Kind, and Martin Peitz. 2012. "Media market concentration, advertising levels, and ad prices." *International Journal of Industrial Organization*, 30(3): 321–325.

Angelucci, Charles, and Julia Cagé. 2019. "Newspapers in times of low advertising revenues." *American Economic Journal: Microeconomics*, 11(3): 319–64.

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

Athey, Susan, Markus M Mobius, and Jenő Pál. 2017. "The impact of aggregators on internet news consumption."

Bakshi, Amar. 2015. "Why and How to Regulate Online Advertising in Online News Publications." *J. MEDIA L. & ETHICS*, 4: 22.

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

Bentz, Tamany Vinson, and Carolina Veltri. 2020. "The Indirect Regulation of Influencer Advertising." *Food and Drug Law Journal*, 75(2): 185–194.

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.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with a continuous treatment." *arXiv preprint arXiv:2107.02637*.

Calvano, Emilio, and Michele Polo. 2020. "Strategic differentiation by business models: Free-to-air and pay-tv." *The Economic Journal*, 130(625): 50–64.

Campbell, Colin, and Pamela E Grimm. 2019. "The challenges native advertising poses: Exploring potential Federal Trade Commission responses and identifying research needs." *Journal of Public Policy & Marketing*, 38(1): 110–123.

Chatterjee, Prabirendra, and Bo Zhou. 2017. "Sponsored Content Advertising in a Two-sided Market."

Depken II, Craig A, and Dennis P Wilson. 2004. "Is advertising a good or a bad? Evidence from US magazine subscriptions." *The Journal of Business*, 77(S2): S61–S80.

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.

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

George, Lisa M, and Joel Waldfogel. 2006. "The New York Times and the market for local newspapers." *American Economic Review*, 96(1): 435–447.

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

Johnson, Benjamin K, Amanda S Bradshaw, Julia Davis, Vanessa Diegue, Lily Frost, Jonathan Hinds, Tracy Lin, Cassidy Mizell, Deanna Quintana, and Ruowen Wang. 2021. "Credible influencers: Sponsored YouTube personalities and effects of warranting cues." *Journal of Media Psychology: Theories, Methods, and Applications*.

Johnson, Garrett, Tesary Lin, James C Cooper, and Liang Zhong. 2023. "COPPAcalypse? The Youtube Settlement's Impact on Kids Content." *The Youtube Settlement's Impact on Kids Content (April 26, 2023)*.

Kerkhof, Anna. 2020. "Advertising and Content Differentiation: Evidence from YouTube."

Mathur, Arunesh, Arvind Narayanan, and Marshini Chetty. 2018. "An empirical study of affiliate marketing disclosures on YouTube and Pinterest." *arXiv preprint arXiv:1803.08488*.

Pattabhiramaiah, Adithya, S Sriram, and Shrihari Sridhar. 2018. "Rising prices under declining preferences: The case of the US print newspaper industry." *Marketing Science*, 37(1): 97–122.

Petrova, Maria. 2011. "Newspapers and Parties: How Advertising Revenues Created an Independent Press." *American Political Science Review*, 105(4): 790–808.

Pfeuffer, Alexander, Xinyu Lu, Yiran Zhang, and Jisu Huh. 2021. "The effect of sponsorship disclosure in YouTube product reviews." *Journal of Current Issues & Research in Advertising*, 42(4): 391–410.

Sahni, Navdeep S, and Harikesh S Nair. 2019. "Sponsorship disclosure and consumer deception: Experimental evidence from native advertising in mobile search." *Marketing Science*.

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.

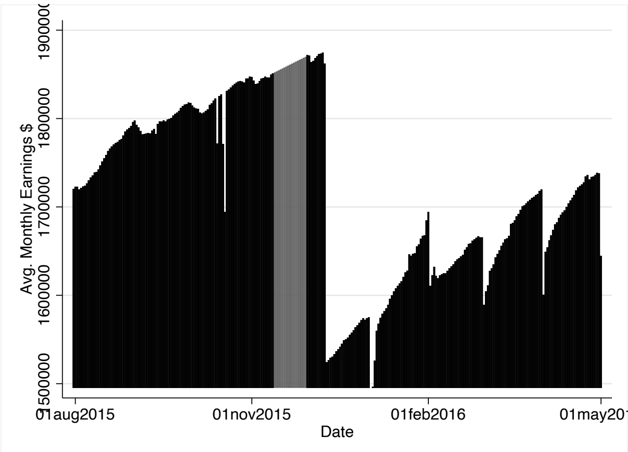
Spence, Michael, and Bruce Owen. 1977. "Television programming, monopolistic competition, and welfare." *The Quarterly Journal of Economics*, 103–126.

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.

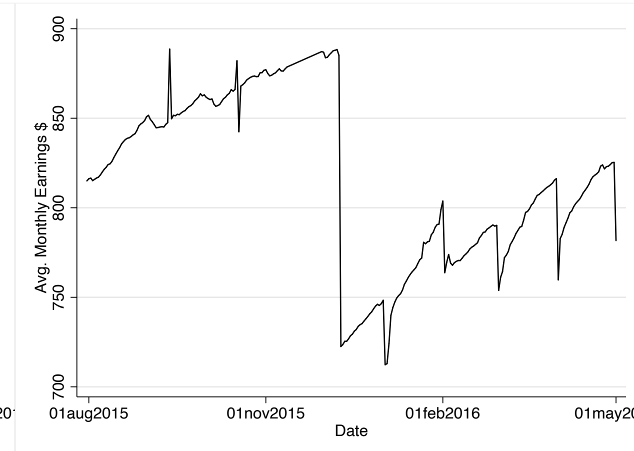
Stennek, Johan. 2014. "Exclusive quality–Why exclusive distribution may benefit the TV-viewers." *Information Economics and Policy*, 26: 42–57.

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

Figures and Tables



(a) Total Monthly Earnings Over Time



(b) Average Monthly Earnings Over Time

Figure 1: Patreon Earnings Before and After Hack

Note: Figure 1a displays total earnings of Patreon video creators. Figure 1b displays average monthly earnings of Patreon video creators. Data on Patreon earnings comes from Graphtreon. Data from 11/12/2015 to 11/30/2015 are imputed due to poor data coverage during that period.

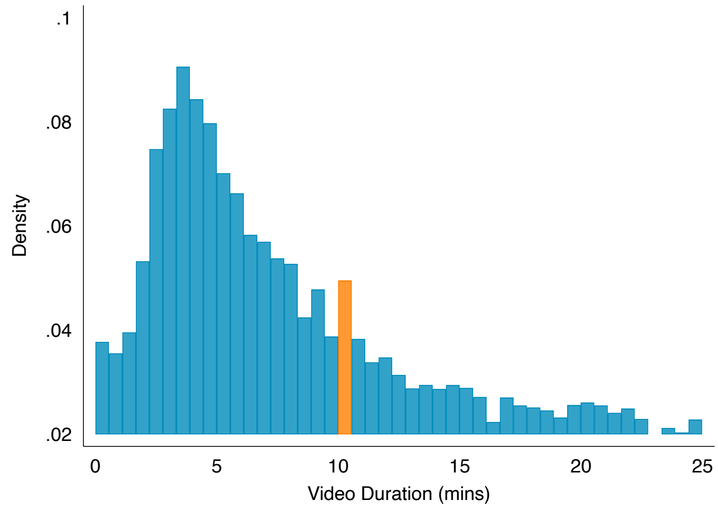


Figure 2: Histogram of the distribution of video duration.

Notes: Sample consists of videos made between August 1, 2015 and July 31, 2016 by Patreon video creators who joined Patreon before May 1, 2015. Data on video duration comes from YouTube.

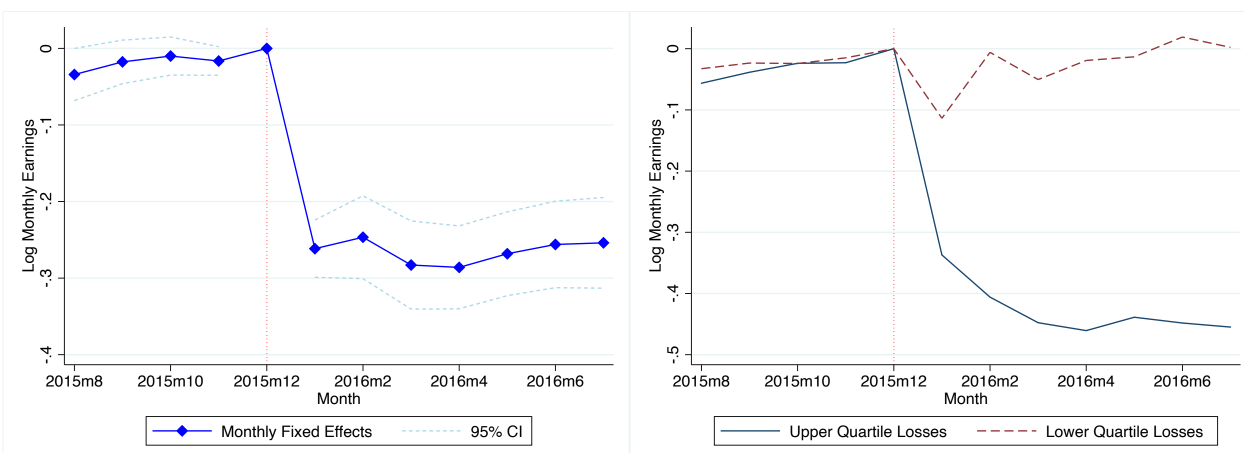
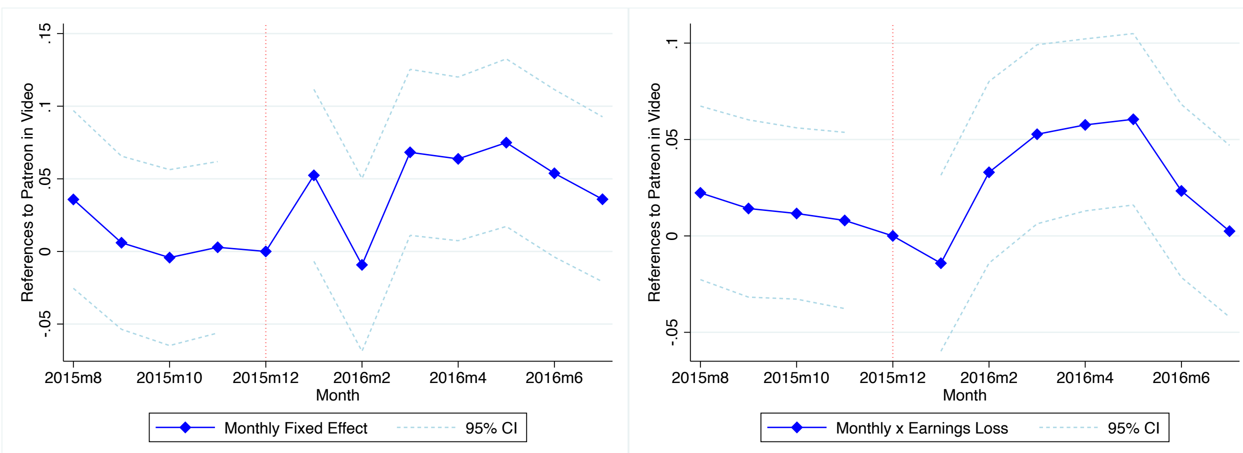


Figure 3: Log Monthly Subscription Revenue Over Time

Notes: Left-hand side presents the coefficients of the month-year dummies from the regression of log earnings on month-year indicators. The right-hand side plots the coefficients separately for creators in the 1st and 4th quartile of losses. Data on Patreon earnings comes from Graphtreon.



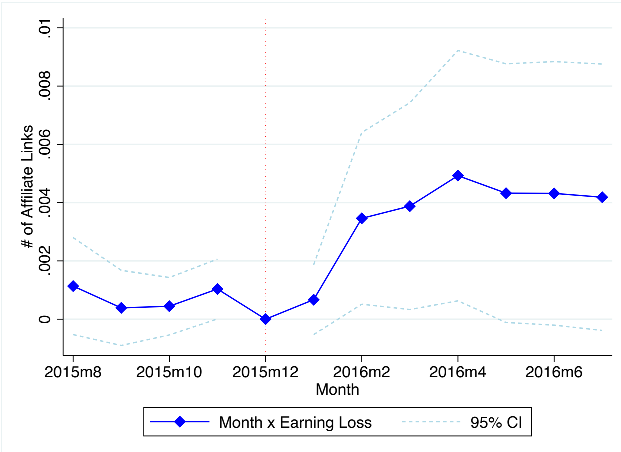
(a) Time-Series

(b) DiD Estimates

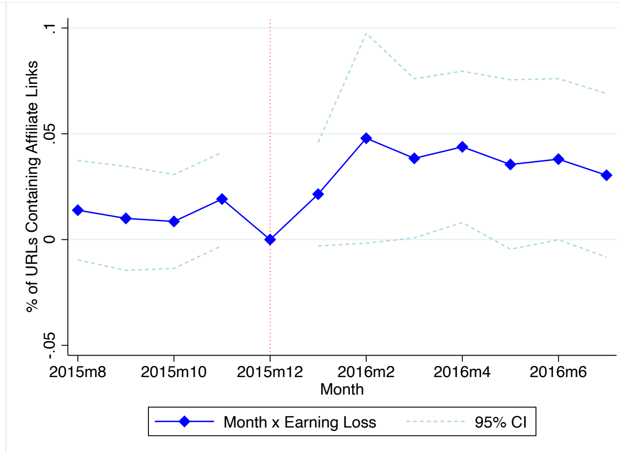
Figure 4: References to Patreon in YouTube Videos Before/After Hack

Notes:

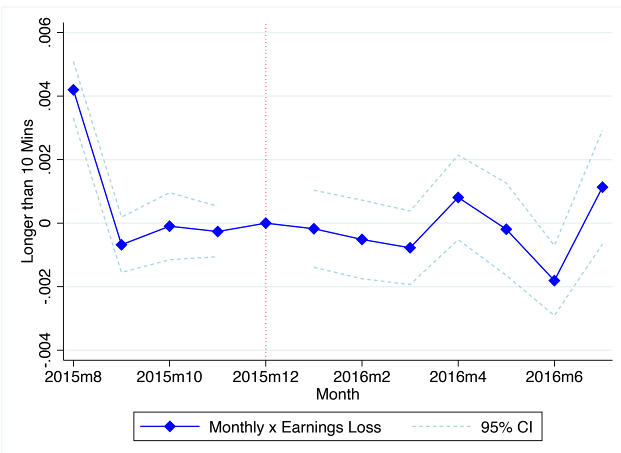
Figure 5: Effects of Subscription Revenue Loss on Advertising Content



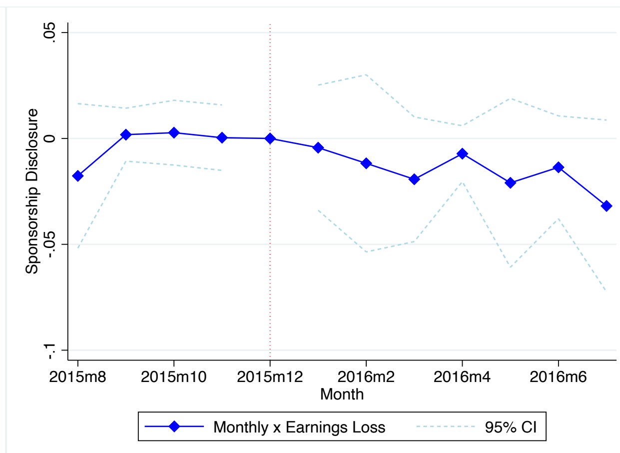
(a) # of Affiliate Links



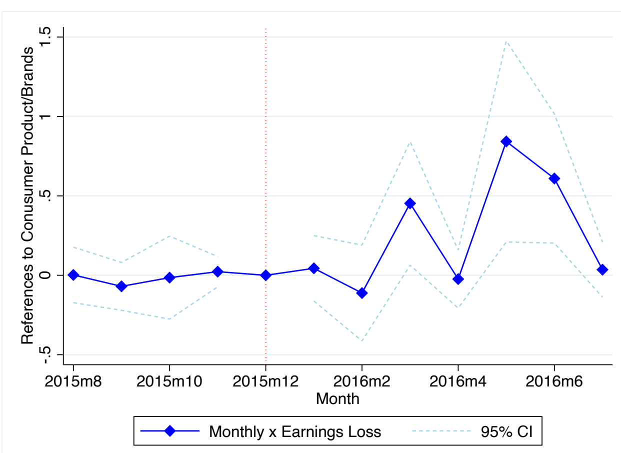
(b) % of URLs Containing Affiliates



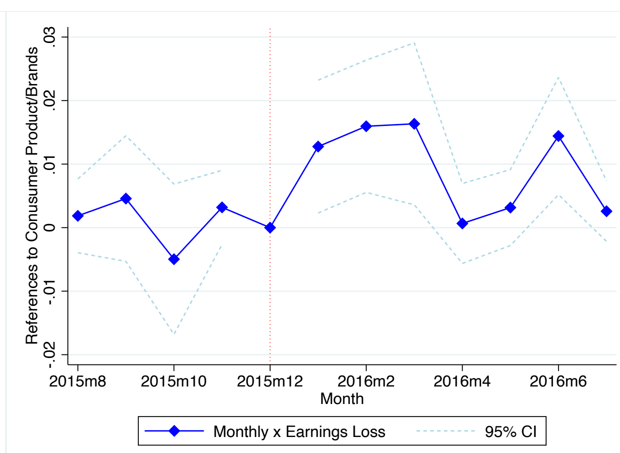
(c) 10 Minute+ Video Duration



(d) Sponsorship Disclosure



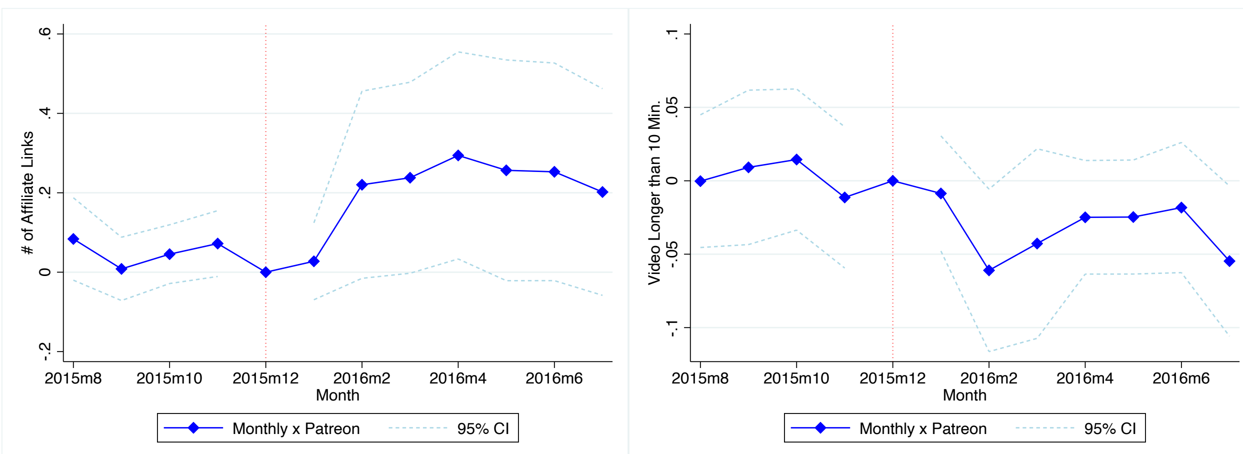
(e) Brand/Product Mentions



(f) Product Reviews

Notes: Figures plot coefficients on the interaction of month-year and earnings loss from dynamic DiD regressions. Standard errors are clustered at YouTuber level. Affiliate links are identified using YouTube video summaries and descriptions. Sponsored disclosures, product reviews, and brand/product and Patreon mentions defined using caption files and video titles. Data on Patreon earnings comes from Graphtreon.

Figure 6: Effects of Subscription Revenue Loss: Early vs. Late Patreon Adopters Specification



(a) # of Affiliate Links

(b) Video Duration Longer Than 10 Mins.

Notes: Sample include both early and late Patreon adopter. See text for detail. Figures plot coefficients on interaction of month-year and early Patreon adoption from dynamic-DiD regressions. Standard errors are clustered at YouTuber level.

Table 1: Descriptive Statistics and Selection into Patreon

	Patreon Early-Adopters				Patreon Late-Adopters			
	Mean	SD	Median	Num Obs	Mean	SD	Median	Num Obs
Patreon Earnings	2335.66	5509.80	779.98	37,984	–	–	–	–
Ad Earnings (From Social Blade)	2710.75	7272.58	196.75	384	–	–	–	–
Likes	2812.88	13128.56	220	35,699	641.11	3985.96	41	139,852
Dislikes	82.86	473.32	10	35,699	53.01	1110.23	1	139,852
Duration (Mins.)	23.56	36.16	12.55	36,246	20.46	35.26	10.77	141,453
Num. of affil. links	0.81	2.08	0	37,984	0.16	1.29	0	148,345
Num. of comments	358.39	1333.28	41	35,526	115.15	934.52	13	139,591
Views (000's)	122.55	658.27	6.87	36,138	32.30	239.38	1.63	141,494

Notes: Patreon "Early-Adopters" are Patreon video creators who joined Patreon before May 1, 2015. Patreon "Late-Adopters" are Patreon video creators who joined Patreon between August 1, 2016 and December 31, 2016. Data comes from YouTube.

Table 2: Effect of Patreon Hack on Subscription Earnings

Dep Var:	Log Earnings		Earnings ($)	
	(1)	(2)	(3)	(4)
Post	-0.2113***	-0.2115***	-140.5395***	-140.5247***
	(0.0087)	(0.0090)	(19.6244)	(20.4913)
Constant	3.3843***	5.4422***	-4303.1749***	950.5456***
	(0.9390)	(0.0052)	(1548.5172)	(11.9513)
Youtuber F.E.	X	X	X	X
Time Trend	X	X	X	X
YouTuber-Specific Time Trend	–	X	–	X
Observations	22337	22337	22339	22339
R^2	0.973	0.989	0.956	0.970

Notes: Each observation is a YouTuber and month-year pair. All specifications include Youtube creator fixed effects. Earnings refer to the monthly earnings on Patreon for each YouTuber at a specific month-year. Data on Patreon earnings comes from Graphtreon. Standard errors are clustered at the YouTuber level. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 3: Effect of Subscription Loss on Affiliate Links and Video Duration

Dep Var:	# of Affiliate Links		% of URLs Affiliate		Video $\geq$ 10 Minutes	
	(1)	(2)	(3)	(4)	(5)	(6)
Post $\times$ Earnings Loss	0.0031**	0.0030**	0.0265*	0.0259*	-0.0012**	-0.0012**
	(0.0015)	(0.0015)	(0.0145)	(0.0140)	(0.0006)	(0.0006)
Post	0.2132**	–	1.6841	–	-0.0155	–
	(0.1012)	–	(1.0990)	–	(0.0261)	–
Constant	0.6943***	0.8147***	6.6591***	7.6100***	0.5788***	0.5701***
	(0.0568)	(0.0004)	(0.6173)	(0.0035)	(0.0146)	(0.0001)
Month-year F.E.	–	X	–	X	–	X
Youtuber F.E.	X	X	X	X	X	X
Observations	37984	37984	37984	37984	36240	36240
R^2	0.830	0.832	0.799	0.800	0.430	0.432

Notes: All regressions are at the video level. Sample includes all videos made between August 1, 2015 and July 31, 2016 by Patreon YouTubers in our sample. Earnings loss is the amount of subscription revenue lost following Patreon hack. "Affiliate links" refers to links in summaries of YouTube videos indicating an affiliate relationship between a Patreon creator and some third party (e.g. Amazon, eBay). Standard errors are clustered at the YouTuber level. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 4: Effect of Subscription Loss on Sponsored Content and Content Commercialization

Dep Var:	Sponsorship Disclosure		Brand/Product Mentions		Product Review	
	(1)	(2)	(3)	(4)	(5)	(6)
Post $\times$ Earnings Loss	-0.0125	-0.0130	0.2822**	0.2832**	0.0080**	0.0083**
	(0.0100)	(0.0101)	(0.1151)	(0.1144)	(0.0034)	(0.0034)
Post	0.0347	–	-0.1279	–	-0.0108*	–
	(0.0226)	–	(0.1194)	–	(0.0064)	–
Constant	0.0308***	0.0523***	0.6704***	0.5917***	0.0127***	0.0059***
	(0.0098)	(0.0048)	(0.0681)	(0.0545)	(0.0037)	(0.0016)
Month-year	–	X	–	X	–	X
Youtuber F.E.	X	X	X	X	X	X
Observations	2235	2235	2235	2235	2235	2235
R^2	0.105	0.111	0.119	0.127	0.207	0.211

Notes: All regressions are at the video level. Sample includes videos with subtitles published by YouTubers in our sample who produced at least one video with subtitles prior and post the hack. Earnings loss is the amount of subscription revenue lost following Patreon hack. Outcomes are constructed using manual-uploaded caption files. In columns (1)-(2), the outcome equals 1 if the video contains explicit disclosure of sponsorship arrangements. In columns (3)-(4), the dependent variable the number of references to consumer brands and products. The outcome in columns (5)-(6) is whether the title contained the word "review". Standard errors are clustered at the YouTuber level. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 5: Effect of Subscription Loss: Alternative Difference-in-Differences

	Num. Affil.	Vid $\geq$ 10 Mins.
	(1)	(2)
Post $\times$ On Patreon	0.1709*	-0.0358
	(0.1015)	(0.0265)
Constant	0.2765***	0.5395***
	(0.0117)	(0.0030)
Observations	186331	177636
R^2	0.612	0.482

Notes: All regressions are at the video level. Sample includes all videos made early and late Patreon adopters (see text for detail) between August 1, 2015 and July 31, 2016. On Patreon is an indicator variable equalling 1 if a YouTuber is an early-Patreon adopter and therefore on Patreon at the time of the hack. Standard errors are clustered at the YouTuber level.* $p < .10$, ** $p < .05$, *** $p < .01$

Table 6: Effect of Subscription Loss: Extensive Margins

Dep Var:	# of Videos		Log (1+ # of Videos)	
	(1)	(2)	(3)	(4)
Post	-0.0483*	0.0000	-0.0048***	0.0000
	(0.0254)	(.)	(0.0014)	(.)
Post $\times$ Earnings Loss	0.0097	0.0101	-0.0010	-0.0010
	(0.0172)	(0.0169)	(0.0014)	(0.0014)
Constant	1.5635***	1.5320***	0.7588***	0.7557***
	(0.0166)	(0.0005)	(0.0009)	(0.0000)
Observations	64908	64908	64908	64908
R^2	0.903	0.903	0.919	0.919
F				

Notes: All regressions are at the YouTuber by month level. Each observation is a YouTuber-month pair. The dependent variable in columns (1) and (2) is the number of YouTube video created by a YouTuber in that month. Column (3) and (4) use the logged number of videos as outcome. Standard errors are clustered at the YouTuber level.* $p < .10$, ** $p < .05$, *** $p < .01$

Table 7: Changes in Subtitle Availability By Treatment Exposure

Dep Var:	Has Manual Subtitle	
	(1)	(2)
Post	0.0575***	–
	(0.0173)	–
Post × Earnings Loss	0.0099	0.0092
	(0.0226)	(0.0225)
Constant	0.3118***	0.3447***
	(0.0098)	(0.0055)
Observations	6451	6451
R^2	0.612	0.613
F		

Notes: All regressions are at the video level. Sample includes videos by YouTubers in our sample who produced at least one video with subtitles prior and post the hack. Earnings loss is the amount of subscription revenue lost following Patreon hack. The outcome is whether the video contained subtitle files. Standard errors are clustered at the YouTuber level. * $p < .10$, ** $p < .05$, *** $p < .01$

A Appendix: Search Phrases

The word basket used to create the word count measure described in Subsubsection 3.2.2 consists of a set of brand names and a set of disclosure statements. In this section, we describe how each of these are constructed.

We obtain from Brandwatch information on consumer brands which regularly employ social-media based marketing campaigns. Brandwatch is a market research company that monitors digital and social media data to to track companies' online presence. Using the firm's "Social Index" rankings, we find the largest online brands across thirteen different consumer sectors. In total, Brandwatch's "Social Index" rankings include 638 different brands. We search for these brands within the captions.

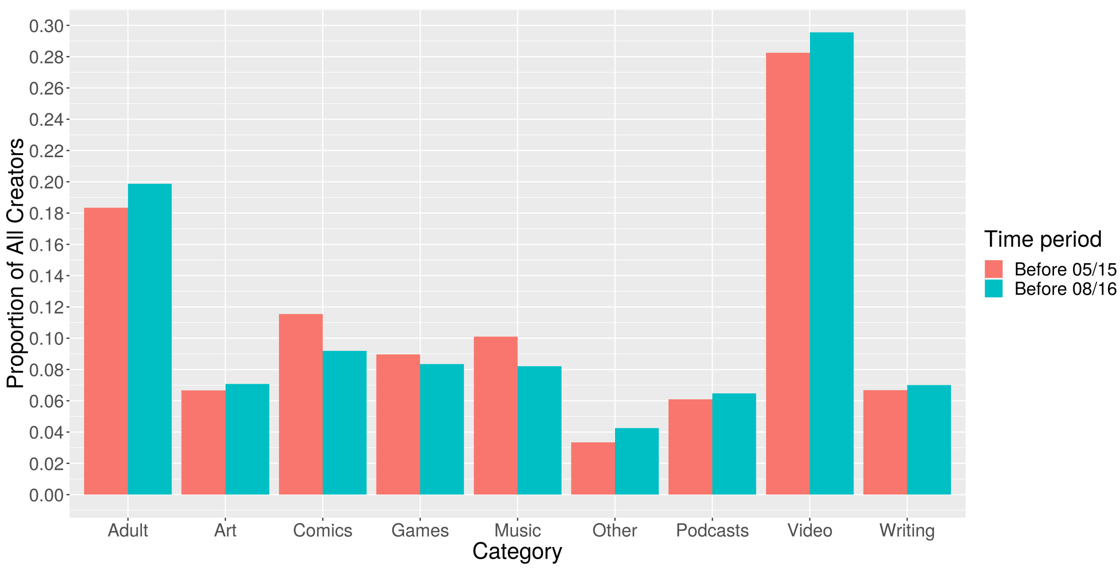
Disclosure statements are phrases YouTubers often use to disclose sponsorship relationships. The disclosure statements we use include:

- video is/was sponsored
- the sponsor(s) of this/my/our/today's video
- sponsoring this/my/our video
- thank my/our sponsor(s)
- the promo/promotional/referral code/link
- this video's/my/our sponsor
- sponsored content
- sponsoring me
- we're/i'm sponsored by
- working with/partnered with/teamed up with

- this video is/was brought to you by
- working/teamed up/partnered/partnering with
- video was made possible by
- collaborating/in collaboration with

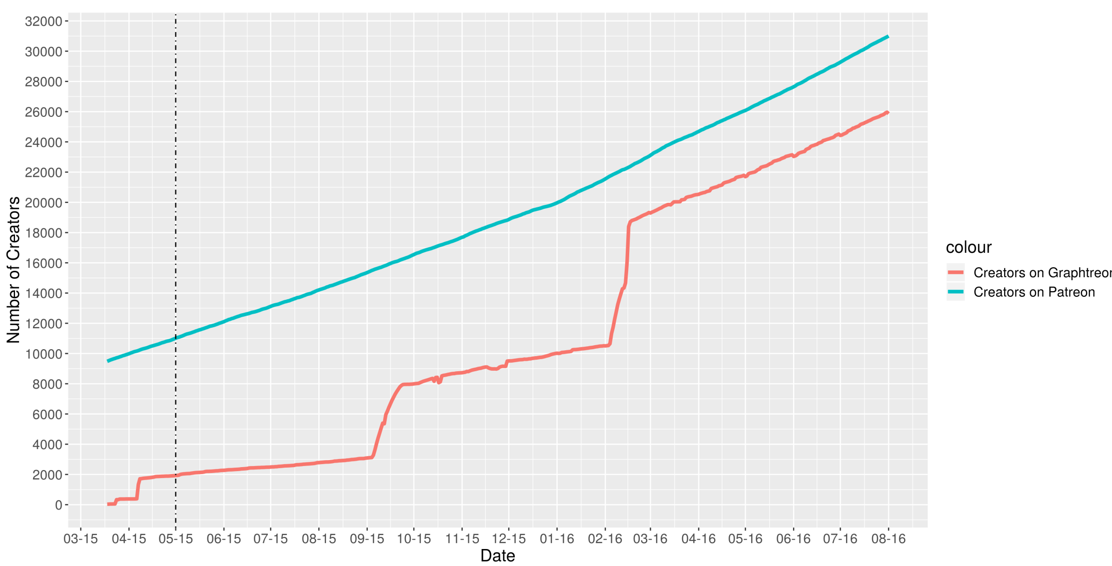
.

Figure A1: Proportions of Creators on Graphtreon by Category



Notes: Data comes from Graphtreon. Time period refers to date at which Patreon creator joined Patreon.

Figure A2: Number of Creators on Graphtreon



Notes: Data is scraped from Graphtreon. Vertical black line displays date by which we require creators to have joined Patreon in order to enter our sample.