Brainrot? The Effect of Short-Form Content on Long-Form Attention*

Adam Di Lizia†

This Version: November 13, 2026

Click here for the latest version

Abstract

Does short-form content impact our attention spans? I study viewing patterns for a large set of news channels on YouTube, using attention heatmaps that record within-video click data. Leveraging the staggered rollout of YouTube Shorts across markets, I show that exposure to short-form content reduces continuous spells of attention to videos. Within-video skipping increases immediately and persists at about 40% above the pre-Shorts mean, with larger effects for longer videos. In turn, I document two important findings. First, the likely mechanism is reduced attention span, since I am able to rule out changing user composition and cross-video substitution as alternative explanations. This is supported by a model in which videos are bundles of interdependent slices where skipping trades off single-slice content against multi-slice context. The effect of short-form content then enters by lowering the value viewers place on context. Second, I show that news viewing patterns on YouTube have changed significantly, with viewers effectively reducing the amount of rich, informational context that they consume. Specifically, a structural model indicates that the value of context in news videos fell by at least a quarter with the introduction of Shorts.

Keywords: Short-form content; attention; news consumption; YouTube; digital platforms.

JEL Codes: L82; L86; D83; D91.

*I am grateful to Manuel Bagues, Renji Bao, Sascha Becker, Christopher Burnitt, Francesco Capozza, Mirko Draca, Ruben Enikolopov, James Fenske, Thiemo Fetzer, Sean Gailmard, Ludovica Gazze, Helios Herrera, Rafael Jimenez-Duran, Sarthak Joshi, Jaime Marques-Pereira, Matthew Ridley, Devesh Rustagi, Lily Shevchenko, Sara Spaziani, Mateusz Stalinski, Gabriel Ulyssea, David Yanagizawa-Drott, Ao Wang, Yanhui Wu, Maiting Zhuang and Ekaterina Zhuravskaya, as well as participants of KCL QPE workshop 2026, Warwick PhD conference, RES 2026 and KCL PPE 2026. I am grateful to the ESRC for helping fund this research agenda.

†Department of Economics, University of Warwick. Email: adam.di-lizia@warwick.ac.uk. Website

1 Introduction

Attention is a scarce resource. Across a wide range of social, political and economic activities, more content is available than could ever be consumed. This issue has become far more salient in light of the recent rise of short-form content. First popularised by *TikTok* in 2018, this content allows users to continually scroll through an endless stream of short videos. *Instagram*, *Facebook* and *YouTube* have now also incorporated short-form content into their wider platform ecosystems.

Unsurprisingly, this type of content has drawn significant criticism for how it potentially inculcates a shortened attention span. This concern is widespread enough that the term "brainrot" was popularised to reflect the suggested causal claim.[1] Brainrot has especially important implications for online news. Online platforms often host news stories alongside large amounts of unrelated content, such as these short videos. Given that a large portion of US adults regularly get their news from video platforms, mainly *YouTube*,[2] exposure to distracting short-form content could substantially alter the attention allocated to news.

However, there is a crucial theoretical ambiguity as to whether "brainrot" occurs in practice. Short-form content might be addictive, and there is evidence for this (Bao 2025), but that does not necessarily imply that consumption of other content is affected. In equilibrium, users most susceptible to the negative impacts of short-form content may avoid consuming it. Users may find some content such as news important, and ensure they allocate all their attention when watching. Alternatively, users may substitute away from news content to consume more short-form videos, but this does not necessarily imply a worsening attention span.

This paper examines the causal link between short-form content and attention spans. I take rich data from news videos on *YouTube*, including *attention heatmaps*. These heatmaps divide each video into 100 equal slices and show the proportion of total clicks (i.e. skipping or replaying) that occur at a given video slice. I explain this measure's construction in detail in Section 4; in my context it acts as a measure of skipping to a certain part of the video. I combine this measure with *YouTube*'s introduction of short-form content, *YouTube Shorts*,[3] in the US in March 2021.

Examining the attention paid across 6 million video slices, I first document a striking change to online news consumption. Since March 2021, there has been a large increase in skipping through non-short videos. This holds across 25 large US news channels, both

[1] See Oxford University Press, "Brain rot" named Oxford Word of the Year 2024. Brainrot is the general concern that ephemeral, forgettable content (including but not limited to short-form content) affects attention span.

[2] See Pew Research Center, "5 facts about Americans and *YouTube*".

[3] *Shorts* sits in its own dedicated tab on the app and website, and hosts videos that are at most 1 minute long (3 minutes since 2024).

traditional and alternative media. The effect is sudden, and sustained for longer videos.

The evidence strongly suggests this is caused by exposure to short-form content. The raw data shows clear mean shifts in the propensity to skip after the introduction of short-form content, and event studies show the effect beginning on the day *Shorts* launched. Leveraging the staggered rollout of *YouTube Shorts*, which was released in India in September 2020, I confirm a causal effect of short-form content on skipping behaviour. The effect size is large, about 40% of the pre-treatment mean. This is not driven by a change in user composition, with no corresponding change in measures of views or engagement. Results are robust to a variety of alternative measures, definitions, strategies and aggregations. *Shorts* has fundamentally impacted how information in news videos is consumed.[4]

However, there are two potential mechanisms which warrant further investigation. One is a "brainrot" story: short-form content exposure erodes the stock of attention, leading to skipping in equilibrium. The alternative mechanism is one of partial substitution: users have limited time so skip through videos to spend more time watching *Shorts*. If the policy does indeed affect attention span, then the effect should be largest for longer videos, which contain the most contextual information. A difference-in-differences across long and short videos within the same channel shows a clear increase in skipping after the introduction of short-form content, revealing that users struggle to watch specifically longer videos. Additional evidence comes from information density:[5] the increase in skipping is largest for the most information-dense videos, with muted effects for the least information-dense videos. Both these results suggest that attention, rather than time, is the binding resource constraint, as users specifically avoid complexity.

Within videos, breaking down the outcome at the individual slice level supports the theory of worsening attention. I find a broadly uniform increase in skipping across all slices of the video except for the very first slice and the last 5% of the video. The null effect on the first slice confirms that the measure is capturing skipping.[6] The null effect on the final part of videos suggests a mechanism of content-seeking behaviour. The end of a video usually serves as a conclusion, or wraps up the video. An impatient user would gain little benefit from these parts of the video compared to the main body. This uniform increase is confirmed by heatmap-level measures of spread, which show that skipping becomes more uniform *within* videos.

Interestingly, on the supply side, channels do not change the existing content they produce. Double machine learning estimates show that channel content does not explain the change in skipping, and word embeddings reveal no change in overall video content

[4]Though the mechanism is unlikely to be news-specific. Whilst it is much harder to abstract from changes in content where transcripts capture less of a video's content, analysis of *MrBeast*, the most subscribed *YouTube* channel, implies this is not unique to news media.

[5]Measured as the words per minute of a given video.

[6]This is the one place in a video users would not need to skip to.

across or within channels. Instead, the channels in my sample begin to additionally produce their own short-form content with a delay. Once they do, the effect on the shortest videos[7] gradually attenuates, as users substitute from the shortest videos to short-form content. This does not occur for long videos, which continue to be skipped more after the policy. The short-form content now produced by these news channels also appears to quickly become more homogeneous, and receives less discussion. Therefore, the channel response, if anything, exacerbates the impact on online news by solidifying short-form content in equilibrium.

A theoretical model of video consumption explains why this result reflects a change in user attention. Users can consume slices of a whole video, gaining utility from a slice of the video on its own (content) and from how that slice relates to a previous part of the video (context). Skipping is an equilibrium response that can occur even without any short-form content as users trade off context for more content. After exposure, short-form content lowers the value of context, so users respond by skipping more in search of exciting content.

I link the average change in skipping to the individual value of content and context. Users may skip through videos that bring little context utility, gaining additional content utility. Therefore, under certain conditions, the choice to skip reveals how *Shorts* on average affects the value users place on video context relative to content. Structural estimation reveals that users value context relative to content in online news videos at least 28% less when exposed to *Shorts*. Following the model, users lose welfare from naive *Shorts* consumption. Rather than consuming *Shorts* only after news, they do so before or during news consumption leading to this attention spillover.

Crucially, this research suggests that there are strong "attention externalities" when users are exposed to short-form content. Users skip uniformly through information-dense, long videos, suggesting a worsening of attention span. I document that this has caused a shift in how news media is consumed online.

Related Literature This paper contributes to five strands of literature. First, it relates to the economics of attention, surveyed comprehensively by Loewenstein and Wojtowicz (2025). I contribute by examining what they term an "attention externality," whereby exposure to one content type damages the capacity to allocate attention elsewhere.[8] This possibility is theorised by Wojtowicz et al. (2019) in a model of habit-forming, attention-maximising behaviour, but to my knowledge this paper provides the first empirical evidence for such a phenomenon. This work sits alongside the rational inattention literature (Maćkowiak and Wiederholt 2009; Matejka and McKay 2015; Sims

[7] When referring to short-form content, I will always use the term "short-form." "Shortest videos" refers to videos that are not *Shorts* but are still short (0 to 10 minutes) in duration.

[8] Through willing consumption rather than a distraction or negative shock.

1998; Sims 2003) and broader models of boundedly rational attention allocation (Gabaix 2014). Here, users are not choosing how precisely to process information in order to make better decisions, but are experiencing a reduction in their capacity to sustain attention for the same information choice.

Second, the paper speaks to the large literature on media diets and their economic and social consequences (DellaVigna and La Ferrara 2015). Within this, it relates specifically to work on social media and platform competition. Allcott et al. (2022) and Braghieri et al. (2022) document that social media is addictive and imposes significant mental health costs, while Aridor (2025) shows that restricting access generates substitution toward similar content. Gylfason (2026) shows users substitute from news websites to *TikTok* when it is available, though here I differ by examining attention within videos on the same platform rather than substitution. For a comprehensive review, see Aridor et al. (2024). Beyond welfare, social media has been shown to shift political behaviour (Ash, Galletta, et al. 2026; Enikolopov et al. 2020). H. Chen and Suen (2023) show theoretically that competition for reader attention across news sites may degrade content quality over time which I find possible evidence for in the increased production of short-form content. I contribute in describing a stark difference in how online media is consumed, specifically through slice level attention, and show that this is credibly driven by exposure to short-form content. In addition, I contribute by examining exposure to a specific part of social media (*Shorts*) rather than the whole, helping understand *how* social media may lead to certain outcomes. Work analysing more broadly the effect of media diets on political outcomes (DellaVigna and Kaplan 2007; Durante et al. 2019; Gentzkow, Jesse M. Shapiro, and Sinkinson 2011; Martin and Yurukoglu 2017) suggests the effect I uncover could have serious downstream consequences.

Third, the paper builds on empirical work showing that attentional failures affect outcomes. A first strand documents that consumers systematically under-react to non-salient information, across financial markets (DellaVigna and Pollet 2009; Hirshleifer and Teoh 2003), taxation (Chetty et al. 2009), consumer finance (Stango and Zinman 2014), and high-stakes goods markets (Lacetera et al. 2012). A second strand shows that distraction or fatigue degrade task performance directly (Coviello et al. 2014; Gibson and Shrader 2018; Sievertsen et al. 2016), with related experimental (Toussaert 2018; Zhong 2024) and observational (Brachet et al. 2012) evidence. I contribute by measuring attention itself, rather than a downstream effect on productivity, and by examining short-form content as the potential "distraction."

Fourth, a related interdisciplinary literature documents that digital stimuli can affect stress, mood, and mental health. (Orben and Przybylski 2019) find a small but consistent negative association between digital technology use and adolescent well-being across three large-scale datasets. (C. A. Kelly and Sharot 2025) show causally that the valence of content browsed shapes mental health in a self-reinforcing loop, and (Beknazar-Yuzbashev

et al. 2025) show that exposure to toxic content on social media reduces welfare even while increasing engagement. I contribute by measuring the reduced-form effect of short-form content on attention directly. I also show how certain digital technologies have impacts on the ability to pay attention to content. Nevertheless these findings suggest a range of psychological channels, such as stress, mood deterioration, or mental health more broadly, may mediate the relationship.

Finally, this paper provides causal evidence on the impact of short-form content on widespread attention. Lab experiments exposing small samples to short videos and immediately measuring outcomes find no consistent effects (Y. Chen et al. 2023; Lin et al. 2024; Mahmoudi et al. 2024; Shanavas et al. 2025), while surveys correlating self-reported usage with self-reported attention difficulties (Phipps and Honeycutt 2025; Puaponpong et al. 2023) face identification challenges. Wei et al. (2026) use MRI machines to examine differences in brain activity when the same content is either split into short videos or presented as one long continuous video. They show that short-form content reduces activity in the parts of the brain that deal with memory formation and meaning processing. This effect is immediate, as shown both from brain scans and in a memory recall test after the exercise. I contribute to the literature using large market level data on video skipping showing that this effect does occur in equilibrium, and leads to a fundamental change in how users watch online news videos through the demand and supply side.

2 Background

This section gives a brief overview of *YouTube* as a social media platform, its history and its introduction of short-form content.

2.1 Overview

YouTube is the largest video sharing platform in the US,[9] and was founded in 2005. Content creators make a channel and can upload a video about any topic. As a result the site hosts an enormous range of content, spanning entertainment, instructional videos, cooking, lifestyle and more. Of particular interest is the range of political and news content which has grown on the site since the early 2010s. *YouTube* videos can be any length, and users can like/dislike[10] the video. Views are measured and displayed to users. If logged in, users can also comment on the video.

Initially, users uploaded videos for the sake of sharing them, but in 2006 *YouTube* began running advertisements. This gave *YouTube* revenue and later on uploaders could earn a share of this ad-revenue depending on how viewed their video was. Given the

[9] See Business of Apps, *YouTube* statistics. Worldwide, *TikTok* has more overall downloads driven by *YouTube* being blocked in China.

[10] Though hidden since 2021, see Freimane (2024)

relatively low barrier of entry and the potentially large returns, *YouTube* has exploded in popularity both for the supply and demand of videos. To give a concrete example, *MrBeast* is the most subscribed *YouTuber*[11] and has an estimated net worth of greater than $2.6 billion (Fortune 2026).

In general, *YouTube* is by far the most used social media platform[12] among adults, and almost 90% of US teens use *YouTube* compared to 60% using *TikTok*.[13] The platform has an incredibly large reach and is important in its own right.

2.2 News Media and Political Content

Political content on the platform varies. Traditional media, such as Fox News or CNN, has been posting regularly since before 2010, and has hundreds of thousands of videos. In recent years, given the low cost of entry, alternative media has risen dramatically. These range from daily news content (e.g. Ben Shapiro, Brian Tyler Cohen) to long form podcasts that may or may not venture into the political (e.g. Joe Rogan, Tucker Carlson, Lex Fridman). The topics of these channels are a subject for further research, but in general they are varied in their political biases and the types of content they produce.

Importantly, *YouTube* is a large supplier of news media to US adults, with 33% saying they regularly get their news from *YouTube*.[14] Changes to how *YouTube* videos are consumed therefore matter for news consumption in the US.

2.3 Short-Form Content

Short-form content was introduced in India in September 2020, in response to that nation's *TikTok* ban. On March 18, 2021, *Shorts* was then introduced in the US and launched world-wide later that year. Short-form content on the platform was limited to 1 minute or less until 2024, and then 3 minutes or less.

An example of a *YouTube* short can be viewed here. A key feature to highlight beyond the length of these videos is the vertical aspect ratio to allow for endless swiping or scrolling.[15] When a user becomes bored with a video, they can simply swipe and the platform feeds them a new short video. Often, *Shorts* embed subtitles on screen or show multiple content streams simultaneously.

To access this content, *YouTube* can recommend short-form videos directly. However, the main method is through a separate *Shorts* tab on mobile and desktop. This is shown

[11] He is additionally popular on other platforms, but his initial rise was driven by exposure within *YouTube*

[12] Whether *YouTube* should be classed as "social media" given the relatively greater asymmetry between producer and consumer is a matter of debate. Nonetheless, the site is used more than other social media platforms.

[13] See Pew Research Center, "Teens, Social Media and Technology 2024".

[14] See Pew Research Center, "5 facts about Americans and *YouTube*".

[15] If on desktop.

in Figure A.1 in the Appendix. *Shorts* provides short, easily skippable, attention grabbing and often disparate content. The nature of *YouTube Shorts* helps motivate why we might expect spillovers on user attention to other videos. To formalise this concern, I now turn to a theoretical model of attention allocation.

3 Model

The purpose of the model is to understand how and why users watch videos. More specifically, it aims to answer why a user may choose to skip a video in equilibrium, and how exposure to *Shorts* might immediately lead to an increase in such behaviour. The model will additionally rationalise later empirical results, and help interpret the observed change in skipping as a change in how individuals value the context of a video.

3.1 Theoretical Model

Consider a video v as a collection of slices s_v which together make up the full video. In this setting users need not watch all of them. They can watch parts of a video instead, either through early exit or skipping. The choice to skip sections of videos will reveal what users pay attention to, both theoretically and empirically.

3.1.1 Set up

Although the data consists of discrete slices, I model consumption as a continuous path to allow for greater generalisability. Users choose a consumption path S_v^* for a given video, so that $S_v \subseteq [0,1]$. For example, $S_v^* = [0,1]$ would be the full video. $S_v^* = [0, 0.1] \cup [0.9, 0.95]$ would represent a user watching the first 10% of a video, then skipping to the final 10% and watching until the final 5%. The crucial takeaway is that videos need not be fully watched or fully unwatched.

In terms of instantaneous benefits from user i watching a given video v slice s, I split the utility into two components as follows:

$$\underbrace{u_v(s)}_{\text{Utility of Slice } s} = \underbrace{q_v(s)}_{\text{Content Utility}} + \underbrace{c_v(s \mid H_i)}_{\text{Context Utility}} \tag{1}$$

Here, $s_v \in [0,1]$. $q(s)$ reflects what I term *content utility*. It depends entirely on the current slice and nothing else, and this utility is the same irrespective of a user's consumption path S_v. I allow it to vary by video, reflecting that videos differ in how interesting they are overall and in *how* this content is distributed.

$c_v(s|H_i)$ reflects what I term *context utility*. This is utility that a user receives if and only if they have consumed previous slices of the video. For example, if the video is making a logical argument, there is some utility gained from watching the concluding

remark provided the previous logical links have been consumed. Here, this is a function of $H_i \subseteq [0,1]$ which is the history of slices watched prior to s. Clearly, the context utility at $s = 0.5$ should vary depending on whether user i has previously watched $H_i = [0, 0.5]$ or $H_i = [0, 0.1]$. Finally, omitted is an outside option normalised to 0. This reflects that there is a time cost to each slice consumed, and so content, context or overall utility may all be negative.

Longer videos still have an $s_v \in [0,1]$ but the context and content utility will vary. They will be higher if 5% of a longer video has more benefit from watching, and lower if that 5% is boring and takes more time relative to the outside option.

General Form of Utility Content utility is fairly straightforward to define. Any sensible mapping $q : [0,1] \to [0,\infty)$ will define the utility from consuming a slice. In the simplest case $q(s) = \hat{q}$, reflecting a constant content utility for the whole video.

For context utility, I assume it accumulates through links between pairs of slices:

$$\underbrace{c(s \mid H_i)}_{\text{Context Utility}} = \int_{\underbrace{s' \in H_i}_{\text{Previous Slices}}} \underbrace{\kappa_v(s'|s)}_{\text{Link value of } s' \text{ and } s} ds' \tag{2}$$

$\kappa_v(s'|s)$ is the value of the link between an earlier slice s' and the current slice s, and is free to vary arbitrarily across and within videos. Context utility at slice s is simply the sum of these links over everything watched so far. This admits a variety of context structures within videos.

The key assumption is that pointwise context utility is derived from aggregating pairwise comparisons. This rules out a payoff requiring three particular slices jointly but nothing if any two were watched. Allowing all arbitrary combinations would mean defining preferences over the power set of $[0,1]$, which is neither tractable in a continuous setting nor a plausible account of how a viewer evaluates a video. For news and long-form discussion, graded accumulation seems a natural description. Nonetheless I impose the pairwise form for tractability and this could be further relaxed.

Note that with this set-up, I assume users know the full utility at any given point. Uncertainty is modelled explicitly in the appendix, but I want to highlight that skipping can occur in equilibrium even abstracting from a possible search or learning channel.[16]

[16]Skipping could additionally arise from users trying to learn about the quality of the video. This extension is considered in Appendix C.

3.1.2 Outcomes

Users choose S_v^* to maximise:

$$\max_{S_v^* \subseteq [0,1]} \underbrace{\int_{S_v^*}}_{\text{Aggregate Over Slices}} \Bigg[\underbrace{q_v(s)}_{\text{Content Utility}} + \underbrace{\int_{H_i(s)} \kappa_v(s'|s)\, ds'}_{\text{Context Utility}} \Bigg] ds \tag{3}$$

Definition 1. *A viewing set $S \subseteq [0,1]$ is consecutive if it is a connected subset of $[0,1]$, i.e. an interval $[a,b]$. A user skips if their optimal viewing set S^* is non-consecutive.*

Proposition 1. *Let S^* be the optimal viewing set under perfect knowledge of $q_v(s)$ and κ_v. S^* is non-consecutive if and only if there exist slices $s_1 < s_2$ with $s_1, s_2 \in S^*$ and $(s_1, s_2) \cap S^* = \emptyset$ such that:*

$$\underbrace{q_v(s_2)}_{\textit{Utility of Skipping to Slice } s_2} > \underbrace{\int_{s_1}^{s_2} \left[q_v(s) + \int_{S^* \cap [0,s]} \kappa_v(s'|s)\, ds' \right] ds}_{\textit{Utility of Watching to Slice } s_2} \tag{4}$$

The content gain from jumping to s_2 exceeds the total utility of watching linearly from s_1 to s_2.

As a result, skipping can occur in equilibrium. This reflects a fundamental trade-off between how users value *content* and *context*. Note that the outside option is included in utility, and so $q_v(s)$ may be negative. This is why the aggregate utility of a section may be lower than the utility at the boundary of a section.

Shorts and Context Formation Now consider a user exposed to some amount of short-form content before watching a given video. Short-form content exposure impairs memory formation and reduces the brain's capacity to encode sequential information.[17] This is consistent with Wei et al. (2026). I capture this effect through a parameter $\phi(e)$, where $\phi(0) = 1$ and $\phi'(e) < 0$. Here, $e \geq 0$ is a scalar reflecting cumulative exposure to short-form content. Context utility becomes:[18]

$$\underbrace{c(s \mid H_i, e)}_{\text{Context Utility}} = \underbrace{\phi(e)}_{\text{Capacity after exposure}} \cdot \underbrace{\int_{H_i(s)} \kappa_v(s'|s)\, ds'}_{\text{Accumulated links}} \tag{5}$$

The effect of prior viewing on context formation is scaled down by $\phi(e)$. Since $\phi'(e) < 0$, greater short-form exposure reduces context utility for any given history H_i. Importantly, this is a property of the user, not the video. The same video offers less contextual value

[17] This can be the direct effect of *Shorts* exposure, or the indirect result of a more fundamental change to some behavior e.g lower patience for example.

[18] Assuming $\phi(0) = 1$ keeps the results in the previous section identical.

to a user with higher e because their capacity to process sequential narrative information is diminished.

Naive Consumption of Shorts Users choose their level of short-form exposure e to maximise utility from *Shorts* consumption $f(e)$ where $f''(e) < 0$. *Shorts* are enjoyable but with diminishing returns and eventually users will always consume a normal video. In the model outside options are included so *Shorts* gives net negative utility after some time. A naive user treats $\phi(e) = \phi_0$ as fixed when choosing e, solving:

$$\max_{e \geq 0} \underbrace{f(e)}_{\text{Utility from } \textit{Shorts}} + \underbrace{U_i(\phi_0)}_{\text{Video utility, capacity taken as fixed}} \tag{6}$$

with first order condition $f'(e^{\text{naive}}) = 0$. This seems realistic; one can imagine a user in a *Shorts* session scrolling until they are bored of *Shorts*. However, the true optimum internalises the cost of short-form exposure on context formation[19]:

$$f'(e^*) = \underbrace{-\phi'(e^*)}_{\text{Context harm rate}} \cdot \underbrace{\int_{S^*} \int_{H_i(s)} \kappa_v(s'|s)\, ds'\, ds}_{\text{Context consumed}} > 0 \tag{7}$$

Since $-\phi'(e^*) > 0$ and context utility is positive, the right hand side is strictly positive. Therefore $f'(e^*) > f'(e^{\text{naive}}) = 0$:

$$e^{\text{naive}} > e^* \tag{8}$$

Naive users overconsume short-form content because they do not internalise the cost that exposure imposes on their capacity for context formation.[20] The wedge $e^{\text{naive}} - e^*$ is increasing in $|\phi'(e^*)|$, i.e. how sensitive memory formation is to short-form exposure, and in the total context utility at stake. Note that a user could simply watch all their *Short* form content last rather than first and the change in the value of context would not bite. Therefore it seems unlikely that users internalise the context cost. Alternatively users are impatient or addictive and cannot easily re-order their consumption. I additionally consider an extension in Appendix C where skipping videos keeps e high, leading to a content-seeking loop under some conditions.

Implications This model suggests that immediate exposure to *Shorts* within a browsing session could change the way users value context in a given video. The significant

[19] If *Shorts* exposure is so damaging then the equality may never hold and $e^* = 0$ is the corner solution. Clearly $e^{\text{naive}} > e^*$ still holds in that setting. Additionally, under some set-ups the second order condition may not fully hold, though the optimum still consumes less *Shorts* than in the naive case. This is characterised in Appendix C.

[20] Alternatively, one can simply assume *Shorts* are uniquely addictive and so even a forward looking agent would over-consume them and lower context value.

prediction is that in equilibrium users readjust their consumption in response, paying less attention by skipping more.

The effect need not be permanent. Exposure to *Shorts* leads to an immediate change that could even completely vanish in a new browsing session. However, in equilibrium users repeatedly over-consume short-form content, and under-consume context. Crucially, *Shorts* exposure leads to an immediate readjustment in the portion of videos watched.

4 Data

This section discusses the main data source, YT-DLP, the channels chosen and how skips are measured. I also provide descriptive statistics cross-sectionally and over time.

4.1 Data Source

This paper combines three distinct pieces of *YouTube* data: video metadata, transcripts, and attention heatmaps. All data comes from YT-DLP,[21] an open source video downloader for *YouTube* videos. I do not directly download the video itself, but otherwise obtain the full metadata of a *YouTube* video. This includes views, likes, comments, length, title and description. This method allows access to data from any publicly available *YouTube* video. Uploaders can split *YouTube* videos into discrete chapters; I know if and exactly when this occurs. All data is at the *video* level (or video-slice level) rather than the individual level. Care will be taken to ensure the effects I find are indeed being driven by changes to individual behaviour.

In addition, I can obtain the transcript of a video where one exists. The transcript of a video can be obtained in a large number of cases (90% of the videos in my sample have a transcript). Either the channel has manually made captions for the video, or since 2021 *YouTube* has used AI to generate a transcript which applies retroactively to all videos on the platform. For an example of a transcript, see Figure B.2 in the Appendix. This partly motivates my choice of news media. Not only are they valuable in their own right, but I need content for which the transcript is a reasonable summary of the entire video. Of course it is imperfect, but it is far more informative to use the transcript of a podcast or a news report than videos with a strong visual component. I use LDA to generate 10 distinct topics for the videos in my sample, and can control for them flexibly, with an example topic shown in Table B.1 in the Appendix.[22]

I also generate text embeddings (Gentzkow, B. Kelly, and Taddy 2019; Ash and Hansen 2023) using the *MiniLM-L6-v2* transformer. Finally, I create TF-IDF text ma-

[21] The GitHub Repository is available here.

[22] This is not the default specification as measurement/sampling error in the transcripts or topic model may result in bias (Ludwig et al. 2025) and because not every video has a downloadable transcript. Nonetheless, both specifications are very similar.

trices to allow me to flexibly control for video content using the standard double-debiased machine learning framework in Chernozhukov et al. (2018). I limit this matrix to the top 5000 terms for computational reasons.

4.2 Attention Heatmaps

The central innovation of this paper is exploiting *YouTube*'s attention heatmap, a slice-level measure of engagement. *YouTube* splits each video into 100 equal slices, regardless of video length, and returns the proportion of users clicking (i.e skipping or replaying) at a given slice. This is then normalised by *YouTube* such that the most clicked section has a value of 1, with all other sections being relative to the most clicked section. The average heatmap is shown in Figure 1. The large spike at the beginning reflects a restart effect as *YouTube* automatically plays the video once clicked. This can easily lead to users missing the very beginning.[23] The other slices have relatively lower values. If there is an overall increase in the proportion of clicks at the rest of the video, this suggests a change in how users watch the video.[24] This can be seen in Figure 2, where I take 5 random heatmaps just before and 5 just after the policy. We can see quite clearly how the heatmap changes, going from relatively flat with one or two spikes to a much higher mean level of skipping and more spikes.[25]

This heatmap data was released in 2022, and is the basis for *YouTube's* "most replayed" or most watched feature. This is displayed after 3 or so days for videos with at least 50,000 views.[26] That this measures skipping, either directly or indirectly, can be seen in Figure 3. The yellow bars highlight a sponsored segment, whereby the creator advertises a product. Afterwards, there is a large mass generated by clicks past the sponsored segment. It seems as though this measures all clicks directly, but if results were driven exclusively by replays then this implies a much stronger effect on user attention.[27] However, given no effect on the very first slice, and results consistent with an impatience channel, it seems the measure picks up a change in skipping.[28]

A natural concern is that the heatmap itself continuously updates in response to new viewing behaviour, which would attenuate my results if it changed around the policy.

[23] And I find no increase at this slice due to the policy, as we might expect.

[24] Technically, this normalisation could attenuate the results if there was an increase in the overall skips for the most watched section. This is unlikely however, as the average heatmap value is very low, and reflects the average propensity to skip. The most active section is usually the beginning or a specific highlight clicked on for different reasons.

[25] The same exercise is repeated for videos greater than 30 minutes, shown in Figure B.1.

[26] I scrape videos without the heatmap and later show there is no change to selection that could be driving the results.

[27] It would completely rule out substitution and require a much stronger effect on attention span.

[28] Even if only replays are counted, skipping could still be measured as in Figure 3 by users who skip forward and then back slightly to the start of a given sentence. The crucial point to stress is that the heatmap data will measure changes in skipping behaviour, and care will be taken to ensure this indeed causes the results I observe.

This does not appear to be the case. The data goes back much further in time than the heatmap's 2022 release. In 2022, there is no change to measured clicking behaviour as shown in Figure B.4 in the Appendix. We would certainly expect that users exposed to this heatmap would click to these highlighted sections in response. This is not just due to the fact that news videos perhaps are unlikely to be rewatched: I take data from *MrBeast*, the largest *YouTube* channel, which is well documented to have incredibly long video tails (Forbes 2024) and see no change in measured clicks either at that time. To summarise, this measure reflects the clicks of users not exposed to the heatmap.

This makes sense on reflection. If the heatmap continually updated, then new users would likely click on the highest spikes, increasing their value and drawing more users towards them. This would result in a useless measure which would be maximised at one point within the video. In contrast, measuring the heatmap based on click behaviour of unexposed users then keeping it fixed would prevent this convergence. As a result, this measure of user clicks is correct at time of upload rather than time of scraping. If this is not the case, this would only attenuate my results.

4.3 Channels

I analyse large *YouTube* channels chosen to span alternative and traditional media as well as the left and right of the political spectrum. These are all US news channels, daily shows or podcasts, with the full list shown in Table 1. Not all of their content is political, but political content features regularly. For completeness, I take videos from the full history of a given channel. However, limits on scraping speed mean I cap this at 10,000 videos per channel, drawing a subsample spread uniformly across a channel's history where more videos exist. Fox News, for example, has around 100,000 videos, so I scrape 1 in 10 across its lifetime. I do not imagine the omitted videos change the results significantly: the effect is driven by a policy external to any individual channel, and is present across channels with both relatively few and relatively many videos.[29]

I additionally scrape short-form content from these channels if it exists. These videos will be excluded from all baseline regressions to avoid a mechanical change in measures around the policy. The channels in my sample do not immediately begin to post *Shorts* and as a result including *Shorts* in the key regressions will not change my results.

4.4 Descriptives

This measurement immediately reveals a striking change to online news consumption since 2021. Figure 4 plots average clicks at a slice over time, and there is an immediate spike at the introduction of *Shorts*. This decays for shorter videos, but is larger and sustained the longer a video is. Videos greater than an hour such as podcasts are consumed completely

[29]The results will be robust to excluding all subsampled channels.

differently since 2021. These are usually less-edited and more detailed and hence we might expect a greater effect if *Shorts* acts through users' attention span.

The one other feature of note is a dip in November 2020, which I speculate reflects less skipping around that specific US election.[30] It disappears in event-studies once within-channel comparison groups are used.

Figure 1 shows where those clicks land within a video. Clicking is concentrated at the very start, and is considerably lower through the middle and the end. Figure 2 gives insight into how this changes due to *Shorts* exposure.

Table 2 reports descriptive statistics for the main variables. Videos vary considerably in length, views, age and information density. The median video is almost 6 minutes long, which is much shorter than the average of around 20 minutes. One caveat is that views and likes are correct *at time of scraping*, and so need not reflect where a video stood at upload. This is unlikely to matter much: news videos are rarely rewatched; there is no visible trend in views that we would expect if older videos kept accumulating watches, and results are robust with or without view controls.

Finally, I use *Google Trends* to document some facts about *YouTube* searches. As a sanity check, Figure A.3 shows that *YouTube* searches for the *YouTube Shorts* topic begin the day of the policy, and stabilise after some time. I also find that there is an increase in *YouTube* searches for "Attention Span" as a topic beginning at the release of *Shorts*. This is shown in Figure A.4. This is purely descriptive, and while it could be users responding to a lower attention span it could be simply users interested in the topic given the release of *Shorts*. Nonetheless the trends suggest that attention span is an important consideration to *YouTube* users after the policy.

5 Empirical Strategy and Results

This section describes the empirical strategy and results. The key question is to more precisely examine the effects of the spike in video skipping around 2021, and examine whether exposure to *Shorts* is the cause of this shift to user attention.

5.1 Across Channels Before and After

The first approach, following on from the raw data plots in Figure 4, is a simple before and after with controls. More formally, channels are treated at varying times throughout their life-cycle. Identification comes from comparing Channel A exposed to *Shorts* three years into its life-cycle to Channel B not exposed to *Shorts* three years into its life-cycle. Clearly, this will not rule out contemporaneous shocks, but will be a good starting point in analysing the magnitude, timing and duration of the effect.

[30] The pattern is not repeated for the previous or later election, nor does it hold across every channel.

Later on, the approach will follow the more canonical differences-in-differences staggered across markets (Sun and Abraham 2021; Callaway and Sant'Anna 2021), and across videos with plausibly varying levels of "exposure" to the introduction of *Shorts* (Callaway, Goodman-Bacon, and Sant'Anna 2024; Acemoglu and Finkelstein 2008).

The following regression is run at the *slice* level[31]:

$$Activity_{sic\tau} = \alpha_{c(d)} + \sigma_{s(d)} + r_\tau + \sum_{\ell=-30,\, \ell\neq -1}^{60} \beta_\ell \mathbf{1}\{\tau - G_c = \ell\} + (\Gamma' X) + \epsilon_{sic\tau} \quad (9)$$

Where $Activity_{sic\tau}$ is the measured proportion of skips at slice s for video i for channel c at relative time τ. $\alpha_{c(d)}$ are channel fixed effects, potentially further interacted with dominant topic fixed effects. $\sigma_{s(d)}$ are slice fixed effects (from 1 to 100) potentially further interacted with dominant topic fixed effects. r_τ are relative time fixed effects. The default specification is at the month level, but I estimate weekly and daily event studies to further examine whether the timing of the change indeed precisely lines up with the policy. $\beta_\ell \mathbf{1}\{\tau - G_c = \ell\}$ is the time since a given channel was treated within its life-cycle and treatment status.[32] $(\Gamma' X)$ are optional controls including polynomials of views, likes, comments or duration and their interaction with channels, slices and time. Additionally, I can control for whether a video, or given slice, has chapters and how many are present. Flexibly controlling for views, likes or comments will assuage concerns that user composition changes due to the policy, but this will be extensively discussed in the Composition paragraph below.

In all cases, given staggered timing within a channel's life-cycle, estimation follows Sun and Abraham (2021).[33]

We could imagine mechanisms by which *Shorts* exposure affects $\Gamma' X$, leading to bad controls. For instance, increased skipping could lead to a lower incidence of liking a video. As a result, all specifications will be estimated with and without controls. Standard errors are clustered at the video level, and robust to additional clustering at the channel[34] or channel $\times$ month level. Video-level clustering is chosen to rule out autocorrelation of skips *within* a given video.

I can further interact channel and slice fixed effects with dominant topic d. In general, dominant topic fixed effects will be excluded by default due to possible expansion bias (Ludwig et al. 2025), but results are robust to their inclusion.[35]

[31] The coefficients are identical if average skips for the video are taken and the regression is run at the video level by the law of iterated expectations.

[32] Alternatively, one can omit relative time fixed effects also and compare before and after means. The results will be similar.

[33] Alternative estimation strategies following Callaway and Sant'Anna (2021) and Borusyak et al. (2024) are shown in Figure B.11.

[34] And robust to additionally using wild cluster bootstrap when clustering at the channel level.

[35] This is mostly to ensure results are not driven by certain news events, but in practice topics have

5.1.1 Results

Figure 5 shows the results of estimating Equation 9. The effect is around 0.07, relative to a pre-policy mean of 0.15. Average skipping at a given slice is therefore about 40% higher. The evidence is consistent with *Shorts* causing a sudden shift in news watching behaviour. Results are robust to alternative clustering, as shown in Figure B.5, and to the addition/ or subtraction of controls (Figure B.9 for additional topic controls, Figure B.8 for no controls). The full set of results varying fixed effects and controls can be found in Table 3.

The event studies show a clear jump at the date *Shorts* is introduced. In some specifications, there are slight trends, though in the opposite direction of the effect. In the main specification with controls or when comparing across videos by length in Section 6.1, pre-trends will be absent.

There is large variation in the effect by video length. I display the effect for videos greater than 30 minutes, but this is a more general pattern regardless of the cut-off chosen, as shown in Figure B.6. For shorter videos[36] the initial effect is similar before decaying after about 7 months. In contrast, the effect for longer videos does decay slightly but remains consistently high for the rest of the sample. This will be explored more thoroughly in Section 7.1, but the fall seems to be driven by a substitution of attention towards short-form content produced by the channels in my sample. This is possible for the shortest videos produced by a channel, but short-form content fails to be an adequate substitute for longer videos resulting in a sustained effect for those videos.

Figure B.10 shows the results of estimating Equation 9 at the daily and weekly level. As we might expect, these estimates are far noisier, but in both cases the average increase begins the *day* that *Shorts* is released. This helps support the overall causal claim and suggests the shift in attention to online news is a result of exposure to short-form content.

5.2 Comparing US and India

Nevertheless, I want to ensure these results are not driven by contemporaneous shocks to user attention, or a platform wide change in behaviour or measurement. For identification, I leverage the staggered rollout of *YouTube Shorts*. *Shorts* was released in India in September 2020, and in the US six months later in March 2021.[37] I take data from 5 Indian *YouTube* channels: The three largest Indian news channels (*AajTak*, *IndiaTV* and *ABP News*) and two large broader media channels (Sandeep Maheshwari and Dhruv

very little effect on average skipping.

[36] Most videos are less than 10 minutes and so the effect is overwhelmingly driven by this type of content.

[37] While I cannot be sure that Indian users do not watch US *Shorts*, such spillovers would only attenuate the results. I find no evidence of this.

Rathee),[38] alongside the same US channels as before.

The first strategy is a simple placebo test. Given the immediacy of the US effect, the *Shorts* effect should be verifiable in the raw data. There should be at the very least no contemporaneous jump for Indian channels, and additionally a jump for Indian channels uniquely in September 2020. Therefore, if *Shorts* causes skipping, we should see two precisely timed mean shifts that only affect their own markets.

I also estimate the following equation which uses under additional assumptions markets as counterfactuals to eachother:

$$Activity_{sict} = \alpha_c + \sigma_s + \sum_{t=-30}^{60} \beta_\ell \mathbf{1}\{t - G_c = \ell\} + (\Gamma' X) + \epsilon_{sict} \tag{10}$$

Where $Activity_{sict}$ is the measured activity at slice s for video i for channel c at time t. α_c and σ_s are channel and slice fixed effects respectively. $\mathbf{1}\{t - G_c = \ell\}$ is a group time indicator relative to treatment status as before. However, now there are two treatment dates (September 2020 and March 2021) for Indian and US channels respectively. $(\Gamma' X)$ are optional controls including polynomials of views, likes and duration and their interaction with channels, slices and time. Standard errors are clustered at the video level (and robust to clustering at the channel level). Estimation follows Sun and Abraham (2021) with an important caveat. After March 2021 both markets are treated, so any specification with calendar-time fixed effects would have to use the already-treated market India as a control, and the two well-timed mean shifts I expect to (and do) find could not be aggregated or identified without some restriction.[39] I therefore omit time fixed effects and compare each market to its own pre-period which allows me to correctly aggregate the two treatment effects. I am essentially assuming that treatment effect heterogeneity does not perfectly cancel out with a common shock at the precise date the US is treated. Figure 7 shows clearly the effects we expect to see if *Shorts* exposure causes skipping, which stagger correctly with treatment dates.

India is of course a substantially different market for news with a different population. The value of the comparison is that the timing of the shock differs between them, so that each checks the other at the moment the other is treated. Additional tests across videos *within* the US market, discussed in Section 6.1, will bolster causal claims in general and with respect to the specific attention-span mechanism.

[38] These channels are more broad than current affairs, and results are robust to their exclusion. However, they are more similar in terms of upload frequency and heatmap availability to the average alternative media channel in my sample.

[39] The issue is that with unrestricted treatment effect dynamics an observed spike in the US 6 months after India is not separable from a sudden drop in the India treatment effect, no treatment effect for the US and a perfectly contemporaneous common shock at the exact date of *Shorts* release. This seems unlikely but cannot be ruled out without a third never treated group. Time fixed effects are then collinear with sustained treatment effects from March 2021.

5.2.1 Results

The effect of *Shorts* can be clearly seen in Figure 7. *Shorts* increases skipping only in India in September 2020. Then, *Shorts* increases skipping only in the US in March 2021. This graph confirms that the jump detected is a feature of *Shorts* exposure rather than a US-specific or platform-specific market shock. This also rules out *YouTube*-wide changes in measurement as there is no change to the Indian channels in my sample at that time. The same two mean shifts are visible restricting the Indian sample to the three news channels alone, as shown in Figure B.13.

This claim is bolstered by Figure 6, which shows the results of estimating Equation 10. These results are very similar to previous estimates. Along with Figure 7, short-form content does seem to causally increase skipping at a given slice of a video. This rules out contemporaneous changes in the algorithm, or *YouTube*'s measurement systems as driving the effect. Taken together with the previous results, there is strong evidence that short-form content shifts how users consume news content, leading to long term changes in attention paid to *YouTube* videos. The treatment effects are shown in Table 4.

In relation to measurement, Figure B.15 shows that this increase is driven by non-consecutive slices rather than a mean shift on every slice. This affirms the results being driven by skipping rather than some measurement change which might presumably affect all slices equally.

Overall, exposure to *Shorts* leads to a significant, immediate increase in skipping across videos. This effect is sustained for longer videos.

The Immediacy of the Effect One stark feature of these results is the immediacy of the effect. Can a jump this sudden really be driven by how users pay attention? Two things must be true: short-form content must have been available in volume immediately, and exposure must have affected behaviour immediately.

First, there was a large amount of short-form content available at the time of release (Violot et al. 2024). This was due to existing videos being re-tagged as *Shorts*, large interest by existing creators and the ability to re-upload *TikTok* videos to *YouTube*.

Second, research shows that *Shorts* could have an immediate effect on how people process information (Wei et al. 2026). It is unlikely that one isolated *Shorts* session will have a permanent impact on a user's watch habits. Rather, one *Shorts* session can have an immediate impact on current activity. We expect this to decay (potentially) quickly but in equilibrium, users continue to consume *Shorts* and this effect is consistently sustained. This type of effect is additionally rationalised by the earlier model, where changes to the value of context are driven by a recent stock of consumed *Shorts* even within a browsing session. Any change will lead to an immediate increase in skipping.

Crucially, if this effect is indeed driven by attention span,[40] it seems users continually expose themselves to *Shorts* which immediately and consistently affects how they process information. As suggested by the model, they respond by skipping through their existing content. Repeated short-term effects are more likely to be reversible, through for example improved blocking options or time limits, compared to a gradual but permanent change. Alternatively, by reordering users consumption, this effect could be mitigated or avoided.[41]

Composition An alternative explanation is a change in viewer composition. My data is at the video-slice level, so if *Shorts* selects low attention span users who proliferate across news channels, then the increase in skipping reflects selection. However, this is not the case for three reasons.

First, as already mentioned, views, likes and comments are flexibly controlled for. Figure B.16 additionally shows no commensurate change in these measures of engagement around the policy. It is difficult to imagine a change in composition that could both explain the sudden and large effect whilst being undetectable.

Second, there is an asymmetry in exposure. Whilst users of the channels in my sample were immediately exposed to short-form content, users of short-form content would *not* have immediately been exposed to the channels in my sample. Figure 15 plots the average proportion of short-form content *in my sample* over time. If the effect is driven by new users discovering these news channels, then there ought to be an increase in skipping when channels in my sample post their own *Shorts*. This is not the case, confirming the effect is on existing users.[42]

Third, *Shorts* grows substantially in popularity over time. In contrast, the effect does not. If *Shorts* is transferring low-attention *Shorts* watchers toward news channels then we would expect the effect to increase as *Shorts* gets more popular. The growth can be seen in Figure A.2. A composition story requires the effect to track the size of the *Shorts* audience, and it does not, which points instead to a change in the behaviour of existing users.

6 Mechanisms

But what is the change in user behaviour causing the effect? This section examines the mechanisms in more detail using variation in video duration and additionally breaking the effect down at the within-video level. The previous section confirms there is a large,

[40] Which I aim to confirm in the following section.

[41] This will be discussed in more detail in Section 6.3.

[42] If anything, there is a decrease in the effect when better substitutes are available which will be discussed in Section 7.1.

causal effect of *Shorts* exposure on skipping through videos. However, there could be different mechanisms that explain the effect.

The first is the attention span mechanism. Short-form content erodes the attention stock of a user. This lowers their value of context through increased boredom or impatience, leading to increased skipping.

Alternatively, the mechanism could be a form of partial substitution. There is a limited amount of time allotted for video content. *Shorts* increases the total available content, and so users substitute to them from the news they watch. This does *not* happen at the video level. If it did, we would see no change in skipping but instead a fall in views of existing videos as users substituted to *Shorts*. Substitution can only be consistent if users skip through existing videos in order to spend more time watching *Shorts* whilst still consuming some amount of their existing video content. Such an effect would be noteworthy: *Shorts* would uniquely produce a *within*-video substitution effect that does not occur with more "long" videos. This would still reflect a fall in the attention paid to news videos, but this would *not* necessarily be due to a fall in attention span.

Finally, supply side responses could mediate the effect. Channels on *YouTube* are monetarily rewarded for engagement and watch time. If channels change their video content in response to either *Shorts* or lower attention, then the observed increase might be a combination of demand and supply factors. This will be examined in Section 7.

6.1 Duration DiD

The most critical implication of the attention span story is a differential effect by video length. Six 10-minute videos and one 60-minute video are identical burdens on time. They are not identical burdens on attention.[43]

Longer videos carry more information and more contextual references, so more has to be held in mind to follow them. In terms of the model, doubling the length would increase contextual references by more than double[44] if the video was referenced throughout. In addition, the longest videos on the channels in my sample are the least edited. These are podcasts and livestreams, where nothing is cut to grab attention. If skipping reflects impatience rather than a shortage of time, it should rise more for long videos. Since the overall effect is immediate, this differential should appear immediately too.

This motivates the following estimating equation:

$$Activity_{sict} = \alpha_{cis(d)} + \sigma_{cts(d)} + \tau_t + \sum_{t=-30}^{60} \beta_t Long_i \times MonthsShorts_t + (\Gamma' X) + \epsilon_{sict} \quad (11)$$

[43]This is partly why *Shorts* is so addictive, following medical arguments from Wei et al. 2026 and allocation arguments from Bao 2025. Splitting content into chunks, whether related or not, changes how we interact with such content.

[44]Assuming each new part has a fixed quality

Every variable is identical to Equation 10, with the exception of $Long_i$ which is a dummy variable equal to 1 if a video is "long". This is defined in most cases as 30, 45 or 60 minutes,[45] and functions as a coarse treatment variable. I will also follow Callaway, Goodman-Bacon, and Sant'Anna (2024) using B-splines to estimate a continuous treatment effect on the treated.[46]

The advantage of the binary treatment is that, while coarse, it allows for event studies to help validate the causal claim. Channel × Month fixed effects will absorb all variation previously examined,[47] and isolate the additional effect between short and long videos within a given channel-month pair. Identification leverages the difference in theoretical exposure to a worsening of attention and controls for changes *within* the US market. If longer videos are skipped through more than short videos, it provides further evidence that overall attention has worsened in response to the policy.[48]

Duration here is not necessarily exogenous with respect to the policy. However, in Section 7 I show that there is no change in duration around the policy for either group. For robustness I estimate the equation using double-debiased machine learning to flexibly control for any changes in video content and the results are similar. Finally, I ensure there is no mechanical effect from slice length changing with video length by optionally controlling for linear duration.

6.1.1 Results

Results of estimating Equation 11 are shown in Figure 8. There is an immediate increase in skipping for longer videos compared to shorter videos after the introduction of *Shorts*. This is of a similar magnitude to the earlier results, with a treatment effect of 0.063 relative to a mean of 0.15. See Table 5 for comparisons across varying fixed effects. Pre-trends hold, and the jump is immediate at the time of the policy.

This is precisely the outcome we would expect given a worsening of attention span. Users pay relatively less attention to longer videos with more contextual information. In the language of the model, longer videos rely more on context utility, increasing the relative price of watching them in full compared to shorter videos.

This effect can also be seen in Figure 9. This admits flexible variation over duration rather than time periods following Callaway, Goodman-Bacon, and Sant'Anna (2024). There is quite a large increase in skipping from 0 to 10 minutes, with a more gradual

[45] The results will be robust to other cut-offs.

[46] One cannot change video length whilst holding all else equal as long videos will be about fundamentally different things, but results will be robust to flexibly controlling for transcript data suggesting the change in the effect is primarily driven by length.

[47] This is important to stress. These results are not a recapitulation of the earlier comparison between videos over 30 minutes and all videos. The channel × month fixed effects absorb that variation entirely.

[48] I see a fall in both, suggesting no spillovers from worsened attention to long videos i.e. through better attention to short videos. However, even if this is the case we still would expect an attenuated difference between the two.

increase as duration grows.

Information Density As stated, a rational forward looking agent who is time constrained has no reason to treat six 10-minute videos differently to one 60-minute video. However, a user might only realise part-way through that they are short of time, which is more likely on very long videos. Relatively more skipping could still reflect substitution whereby time constrained users skip through longer news videos in order to watch more *Shorts*.[49] It is difficult to square this with Figure 9, as users still skip through shorter videos and the marginal increase is large even within the first 10 minutes. The ideal test is one where the context or density of a video increases, but time is held fixed. This would have clear implications.

If time is constrained, then the relative price of length should be driven by the least information dense videos. For example, a user who has only 20 minutes to watch long videos should skip more aggressively through videos that are slower and contain less information. These are precisely the videos that these users wish to avoid if time is short.[50] However, if attention is constrained, we should expect to see the reverse. The increase in skipping should be driven by the most information dense videos. A relatively easy-to-follow video will have a lower attention cost than one with lots of information in a short time, and so skipping should be higher.

This logic builds on from the model. Information dense videos have large amounts of context utility in a shorter period of time. In contrast, information sparse videos have little context utility in a longer period of time. An attention constrained user therefore skips more through information dense videos, as the highly present context becomes less valuable. A time constrained user has a higher outside option (short-form content) so values context utility the same. They would skip relatively more through information sparse videos which take longer to reach the same context or content utility. One is driven by a substitution effect (context less valuable) and the other is an income effect (both less valuable due to higher outside option).

Using words per minute as a proxy for information density, I split the sample at the median and estimate Equation 11 separately within each half. I compare long against short videos, before and after *Shorts*, among videos that deliver information at a similar rate. This keeps time fixed but varies information density (i.e. context links). As we can see in Figure 10, the relative effect is driven by skipping on the most information dense videos. This again affirms an attention span story.[51]

[49] In the language of the model, this would be that preferences are not homothetic with respect to time.

[50] The same would be true if skipping was beneficial because users are better at searching for information since exposure to *Shorts*. If this is the case, then we would again see more skipping for the least dense videos.

[51] It also ensures that differences between slices lengths for longer videos do not drive the results. Comparing less dense long videos to short videos before and after finds no clear effect suggesting this is not mechanical.

6.2 Within Videos

I have established that the average level of skipping increases, but not how this increase is distributed *within* videos. A rise in the mean is equally consistent with uniform skipping through a video, cutting it into sections, skipping only to early slices and then exiting the video, or jumping straight to the conclusion. These imply quite different things about the viewer, and each would leave a different signature in the heatmap.

Attention and partial substitution mechanisms also make opposing predictions following the model. A time constraint changes how much of a video a user can watch, but not what any part of it is worth. Context remains valuable so we would expect videos to be split into sections. News videos make this especially easy, since they are often a sequence of separate stories.[52] Skipping would rise on average but become *more* concentrated, since the cuts fall in particular places. A worsening attention span works differently because it lowers the value of context more generally. Skipping should rise on average while becoming *less* concentrated, appearing throughout the video rather than at a few breaks. The slice-level results and measures of spread below favour the second story.

6.2.1 Slice Level Results

Equation 9 can be re-estimated slice by slice, returning a separate treatment effect at each of the 100 positions in a video. Figure 11 shows the result. Skipping rises by a similar amount almost everywhere, with no tendency to grow or fade as slices progress.[53] This is what the attention span story predicts. Were time the binding constraint, users would cut at particular points,[54] dropping parts of the video and leaving the rest intact. Instead the increase is continuous across almost every slice suggesting an overall lower attention span.

Two exceptions stand out. There is no increase at slice 1, exactly what we would expect since the very beginning is the one place a user never needs to skip to. There is also a sharp drop-off in the final 5% of a video. This is a null effect, as seen in the event studies.

The end of a *YouTube* news video differs from the body. It is often where the presenter asks viewers to "like and subscribe" or signposts future videos, and it is where the video summarises or concludes. Neither appeals to a content seeking viewer: the sign-off contains no new information and the conclusion is unlikely to contain new content.

[52] This will be explicitly analysed by looking at the behaviour when videos contain separate chapters.

[53] Event studies for each individual slice are shown in Figures B.24–B.27 in the Appendix. They corroborate that this pattern is driven by exposure to *Shorts*.

[54] Even if we assume this is still substitution, the observed pattern means users have chosen functionally to give up context for the sake of time, when they could watch fewer videos or cut the video at specific points.

This aligns with the model in suggesting a level of rationality in how users respond. Pure habit formation would result in a uniform increase across the *whole* video. However, there is a clear (relative) distaste for skipping to the very end of videos. This suggests content seeking as the likely mechanism driving these results. Users' attention span has fallen, but they do not stop watching news. They still watch news videos, but skip through them more. They avoid skipping to the parts of the videos that would not have novel information, or require watching the full video to understand. They are not looking for a summary, otherwise they could skip exclusively to the end and would do so more on the least information dense videos. Instead, they are looking for content within these videos that catches their attention, likely driven by *Shorts* eroding their value of context.

This variation between the body and the end can be leveraged to estimate another difference-in-differences, comparing the body of videos to the end before and after the introduction of *Shorts*. Video fixed effects can be included, and the results in Figure B.21 help bolster the causal claim.

6.2.2 Effects on Skipping Distribution

Given that each video returns 100 ordered heatmap values, measures beyond the mean can be calculated for each video. A change in the average of such measures (for example the average median, or the average variance) can be consistently estimated.[55] These are needed because the slice-level results are not quite decisive on their own. Users could be cutting each video at a few key points, with those points falling in different places from video to video, which would still average out to a uniform increase. Measures of spread distinguish the two, since a genuine uniform increase raises within-video dispersion while cutting at a few points will not.

I confirm this using four complementary measures of within-video dispersion: the Gini coefficient, entropy, higher moments (skewness and kurtosis), and jaggedness. The Gini coefficient measures the inequality within the heatmap. If skipping really is becoming more uniform, the average Gini coefficient of the heatmap should fall as skipping becomes more evenly distributed across slices. This is indeed what I find, as shown in Figure 12. I additionally examine entropy, a measure of how uniform a distribution is.[56] If skipping becomes more uniform, the average entropy of a given heatmap should increase. This is also what I find, as shown in Figure 13.

Average skewness and average kurtosis also fall, as shown in Figure B.22. This is consistent with a uniform increase in skipping, rather than a substitution style story. If certain slices became cut points there would be a very large amount of skipping at those points, leading to a skewed distribution with heavy tails.

[55] Following the continuous mapping theorem (Billingsley 2012).

[56] The negative KL divergence between a given distribution and the uniform.

However, these results do not imply that all videos see a mean shift at every slice. Within a video I measure the absolute deviation[57] between two slices. This functions as a measure of jaggedness, the difference between consecutive slices. We see a significant increase as shown in Figure B.23. The increase is *more* uniform but there is a greater change in the rate of skipping between slices. This suggests a mechanism where users skip to certain highlight points within a video. When they do skip, it is far enough to land beyond the next slice, suggesting some users skip large amounts of context.

6.2.3 Chapters

Some *YouTube* videos are split into chapters by their uploader, and these are marked on the progress bar. A user can see when a new chapter begins, and if bored may skip straight to it. Since I observe the precise location of every chapter break, I can build a dummy for whether a given slice contains one. Table B.2 reports the effect of that dummy, both overall and allowing it to vary after *Shorts*. A chapter marker raises skipping by roughly a quarter as much as the policy before *Shorts*, and about as much as the policy after *Shorts*.

These magnitudes give a useful benchmark for the size of the *Shorts* effect. Exposure to short-form content raises skipping more than marking every slice in the video as the start of a new chapter.[58] This is a natural result from a content seeking mechanism, where *Shorts* lowers the value of context and each skip becomes a search for something new.

Figure B.28 estimates a difference-in-differences comparing slices with chapter markers to those without, before and after *Shorts*, and the gap widens. The model accounts for this directly. A chapter start is where a new story or topic begins, so it carries little or no accumulated context. When the value of context falls, slices that never depended on it in the first place become relatively more attractive destinations, and a content seeking user goes there in preference to points that require having followed what came before.[59]

This is difficult to reconcile with substitution. A time constrained user could use chapters to break a video into chunks, but that account already sits badly with the greater spread of skipping and with the effect being concentrated in the most information dense videos. Panel B of Figure B.28 shows the results of estimating equation 9 on chaptered videos while dropping the chapter slices themselves. If chapters begin at slices 10 and 20, those two are omitted and everything between them is kept. Skipping still rises. Users are skipping *within* chapters, not merely to them. Had they been slicing long videos

[57] The absolute deviation is taken to prevent the sum telescoping and cancelling out slice level deviations.

[58] This is a back of the envelope calculation, as of course one imagines that more chapters would necessarily be in competition with each other for attention.

[59] A search and learning mechanism would imply the same, with users skipping to a place they know new content begins.

into manageable pieces, with the video itself marking where they are, there would be no reason for skipping inside those pieces to increase. Alongside the evidence in the rest of this section, the results strongly suggest a fall in user attention span.

6.3 Structural Estimation

The results so far establish that skipping rose and describe where within a video it rose. This section prices that change in content and context terms leveraging the theoretical model in Section 3, and examines which sections of the video benefit due to this change.

Following the earlier model, the utility of skipping[60] is given by the form:

$$\underbrace{\tilde{u}_{isv}}_{\text{Utility of Skipping}} = \underbrace{\Delta q_{sv}}_{\text{Content Gain}} - \underbrace{\Delta c_{sv}}_{\text{Context Lost}} \tag{12}$$

This reflects that the utility of skipping is a trade-off between the net content gained and the context lost. It will vary across videos and across users, as the skipping decision is a function of the video content and the specific individual. The object of interest here is the overall change in the value of context after *Shorts* exposure.

I do not observe individual consumption bundles, only the share of skips aggregated across viewers and slices. In order to begin modelling the effect on context and content, it is helpful to consider two extremes. Consider the value of context to content to be 0. That is, context has no value to viewers. If skipping were blind, a viewer would aim to skip as much as possible to search for valuable content resulting in an average activity approaching share of 1 across the whole video. In contrast, if the ratio of context to content approached infinity then there would be in the limit no skipping whatsoever.[61] Users gain no benefit content seeking and watch the full video to maximise utility from pairwise contextual links.

Of course, neither assumption is likely to hold completely, but 0 to 1 is the absolute largest differential between skipping outcomes. This will give a lower bound on the fall of the value in context.

This still leaves the shape in between unpinned, and skipping at 0.5 is not necessarily the point at which content and context are equally valuable. Context accumulates continuously over a history while content is realised once at the destination, so there is no reason equal per-unit value should correspond to equal skipping odds, or to Activity of 0.5. I take either the pre-period mean of skipping as the index, and price the fall in context relative to the context value before *Shorts*. Alternatively, I take 0.5 as the lower bound of the fall in context.

Equation 12 says a viewer skips whenever her content gain exceeds the context she

[60] Derived fully in Appendix D.

[61] Provided context always has some value such that users would not skip entire sections. The proof is again formalised in Appendix D.

gives up, $\Delta q_{sv} > \Delta c_s$. $\Delta c_s \equiv \phi_t D_s$ prices a time-invariant context-stock gap D_s at ϕ_t, the current value of context affected by *Shorts* exposure. Activity in period t is therefore the population share of opportunities where:

$$a_t = \Pr\left(\Delta q_{sv} > \phi_t D_s\right). \tag{13}$$

Appendix D shows that if Δq_{sv} and D_s are independent and exponentially distributed across opportunities, (13) collapses to

$$a_t = \frac{1}{1 + \kappa \phi_t}, \qquad \kappa \equiv \frac{1 - a_0}{a_0}, \tag{14}$$

a logistic function of $\ln \phi_t$,. Intuitively, with a mass of users and each skipping decision being a function of the value of context and time-invariant video characteristics, the change in proportion of skips tells us about the change in the value of context. This then allows the observed mean shift in skipping to be translated into a lower-bound fall in context relative to content after the policy.

6.3.1 Results

Activity at a given slice rose from a pre-policy mean of $a_0 = 0.15$ by 0.08 (Figure 5), to $a_1 = 0.23$. By (14) this identifies $\phi_1 \approx 0.59$, a 41% fall in the value of context relative to content benchmarked at the pre-period. Appendix D gives a further conservative estimate by assuming that the value between content and context being equal is given when average skipping is 0.5.[62] This gives a lower bound that requires no pre-period normalisation at a 28% fall in the value of context relative to content. This requires no anchoring to the pre-period mean.

It becomes harder to derive utility from context, so users value content more highly and increasingly consume disjoint subsets of a video. This content-seeking behaviour lowers the total context consumed from news videos, of potential concern to the individual viewer and to policy makers who may desire news viewers to consume context. I discuss what this might mean for individual welfare in Section 8.3.

6.3.2 Destination Choices

Conditional on skipping at all, I then aim to examine where users skip. This has important implications in understanding which parts of news videos are relatively more successful after the policy. This differs to the previous exercise within videos, which showed that each slice except the final 5% saw an absolute increase from the policy. Conditional on this increase, do the *relative* choices of slices change after the policy? This is hinted at

[62] Which is extremely large, but given context is weakly additive and content is not this is the upper bound for an index

by the earlier results suggesting that chapter headings benefit disproportionately more from the policy.

I continue to make the assumption that skips are blind. The stylised model allowed for users to know precisely the value at any point in a video, but in truth we would expect users to be completely unsure of the value at a random point in the video. If this were the only consideration, then we would expect no change in the relative skipping destinations. However, before skipping, users will observe chapter headings on the progress bar of a video. This is important, as chapters likely have no contextual links, or far fewer contextual links, relative to other parts of the video. These chapters often start a new topic or news story and so if the value of context has fallen these skip destinations are relatively more attractive. This follows a simple relative price logic derived in Appendix D.

I estimate a model which normalises skipping at chapters to 0, and estimates the relative rate of skipping compared to chapter starts before and after the policy change.[63] I examine absolute time rather than slices to account for longer videos having more context build-up between slices following the model. Figure 14 shows the change in the relative importance of slices after the policy. As we can see, chapter starts become more important, and so do sections of the video near where a chapter starts. This decays to a broadly flat gradient after around 3 minutes.

For a further structural interpretation, if we follow the stylised model in Appendix D we can interpret the shape from the chapter as the context build up in news videos. Intuitively, it is identified by the shift in the value of context from the policy. If context completely resets at a new chapter, and builds up according to some simple function, then Figure 14 identifies such a function. Of course, these are strong assumptions. More broadly, the shape of the graph regardless tells us that context is lowest at a new chapter, and how it builds up as time since a new chapter progresses. News videos in general seem to build up contextual links quite quickly, and suggests that attention is relatively diverted towards the start of stories or arguments.

Shortly after a new chapter start, context has not built up enough to completely devalue skipping there, leading to a relative benefit from the start of discussions.[64]However, after 5 minutes there is only a small marginal loss to skipping further away from a new chapter, as at this point the video is likely making many contextual references regardless. Users will still skip through here as shown in Figure 11, but not differentially so.

Overall, these results suggest that context became less valuable. This led to an increase in skipping everywhere, and a relative shift to low-context chapter starts. We

[63]Estimated on the slice-level heatmap, restricted to chaptered videos, with video and uploader×slice fixed effects.

[64]This does not contradict the earlier results suggesting a uniform increase across slices. However, netting this increase out and examining videos with chapters reveals that there is still an additional difference between video slices at chapter starts.

additionally can see the broad shape of how context builds up in news videos on *YouTube* if we assume no change in content after the policy, which will be examined in the following section.

7 Alternative Explanations

The results so far point to a change in how users watch videos. I next consider alternative explanations that could produce or affect these results.

7.1 Supply Side Effects

Do channels respond to the policy? There are two possibilities to consider. The first is a direct response to *Shorts*. Channels observe the existence of short-form content, and could revise their own videos to be better substitutes or complements for it. This could mediate the observed change in attention, since the videos themselves have changed, leading to increased or decreased[65] skipping.

The second is that *Shorts* does indeed have an effect on attention span, but channels react to mitigate the outcomes. This is plausible as *YouTube* rewards channels for watch time and engagement. This could lead to observed skipping being an underestimate of the true effect, as channels try to make their long-form content more attention grabbing. Channels could also make their own *Shorts* in order to capitalise on the new market structure. The following sections test whether channels change their content directly, and whether they respond by shifting into short-form production themselves.

7.1.1 Changes in Video Content

First, we can examine the transcript of channels' long videos around the policy. Using text embeddings, we can compare usual news videos and observe whether there is any immediate or gradual change in video content. Text embeddings project the transcript into a high dimensional vector space, and are a good measure of video content used elsewhere in economics (Gentzkow, B. Kelly, and Taddy 2019; Ash and Hansen 2023).

Figure 17 shows the change in two measures. The first compares the average cosine similarity from a video to the previous week's embedding. The second does the same for the current week's embedding. We can see that videos do not suddenly become more different within or across weeks either immediately or gradually over time.

Finally, I take the average first principal component of video embeddings as a broad measure to abstract from choice of comparisons. There does not seem to be a mean shift in video content measured with embeddings.

[65] To the counterfactual with no change in videos.

Double machine learning bolsters this claim. Text embeddings may not pick up slight changes in language which do not fundamentally shift meaning but could shift attention. However, if the presence of even one or two new words explains the change in attention, double ML will account for that and return a null result. Table 6 additionally rules out that language changes in general could drive the effect.[66]

Overall, there is no evidence that video content changed in response to the policy, nor that such changes explain the rise in skipping. These measures aggregate over features of the text and could miss very subtle shifts in language. But a shift subtle enough to escape all of them is unlikely to also explain the very large increase in skipping.

7.1.2 Entry into Short-Form Production

One outcome of the policy is the ability for the channels in my sample to create their own *Shorts*. Interestingly, these channels do not immediately do so, taking months before short-form content starts becoming a substantial share of video output. This delay aligns with descriptives in Violot et al. (2024) showing little political or media content in the initial stock of *YouTube Shorts*.

Nonetheless, these channels do eventually begin producing their own short-form content. As shown in Figure 15 there is a discrete jump after 7 months, from which the share continues to tick upward over time.

Recall the earlier fall in average skipping for the shortest videos. In Figure 15, overlaying the time series of skipping and the proportion of *Shorts* reveals the decrease in the effect begins precisely when these channels begin producing their own short-form content. Furthermore, flexibly controlling for the proportion of *Shorts* produced in a given week in Equation 9 results in an overall consistent effect on the shortest videos as shown in Figure B.32.

The pattern is consistent with substitution. Exposed users would prefer to consume shorter videos as *Shorts*. For example, I may not be willing to substitute *Ben Shapiro's* videos for a generic short-form video, but I may prefer to consume *Ben Shapiro's* videos as *Shorts*. Until these channels produce their own *Shorts* this is impossible; once they do, users substitute directly from the shorter usual videos to short-form content.

The lack of a reduction for the longest videos supports this. A 5-minute news video might be possible to consume as one or more *Shorts*, but a 1 hour or more podcast episode will be more difficult to substitute with short-form content. As a result, the effect decays for the shortest usual videos, but is sustained for videos 20 minutes or longer.

This need not show up as fewer views. In contrast to the immediate jump at release, the change is gradual enough that I might not pick up such a clear effect. Alternatively, users may still click on these videos, but have already been exposed to the related news

[66] This is slightly different to the previous point. The results could still not change for double ML, and channels could still attempt unsuccessfully to change their video content.

story or video content in their earlier *Shorts* session and so directly click off the video rather than skip through.

7.1.3 Short-Form Video Features

The *Shorts* produced by these channels also seem to converge. Figure B.29 plots the difference between the cosine similarity from a *Short* to the parent channel and the cosine similarity from a *Short* to other *Shorts*. This measures if *Shorts* are more similar to their parent channel than other *Shorts*. Over time *Shorts* become as similar[67] to each other as to their parent channels. This suggests *Shorts* become potentially more homogeneous over time.

Additionally, these *Shorts* receive more likes per view, but significantly fewer comments per view. This suggests the degree of engagement with news content is falling. Finally, *Shorts* have a higher word-per-minute and more sentences per minute, suggesting uploaders try to compress information into the 1 minute time limit. These can be seen in Figure B.30.

Crucially, the supply side response exacerbates the effect. Whilst the skipping seems to fall, this is not driven by content becoming more interesting to compensate. Rather, the most affected users move fully to *Shorts* to consume news media content and continue to skip through longer videos.

7.2 Sample Selection

A concern could be that after *Shorts*, there is some change to which videos get heatmap data. Either different videos now pass the 50,000 views threshold after the policy, or *YouTube* coincidentally changes which videos have heatmap data. This could result in measuring different videos that had no heatmap data before, increasing measured average skipping through selection. This seems implausible given how little video content explains the results, but can be checked given non-heatmap videos are also scraped.

Figure B.34 shows that the probability for a video to have a heatmap does not change around the policy, or differentially between longer and shorter videos. In addition, the margin driving selection (views) is controlled for and does not vary around the policy as already discussed. Therefore, selection would require a precise balance between videos entering and exiting.

To further rule this channel out, I use double debiased machine learning (DDML) in the presence of sample selection (Bia et al. 2024). Machine learning methods are used to flexibly predict either the probability of selection based on text data, or the predicted skipping behaviour using text data. The outcome when selection is false is not observed. It is doubly robust as if the propensity score model is well specified, selection

[67] At least according to the embedding space.

can be predicted. If the outcome model is well specified, activity can be imputed. In practice, neither will be exactly correct but the method uses both to minimise the bias from selection.

The results are shown in Table 6. DDML on sample selection results in a similar estimated treatment coefficient. Crucially, the selection of which videos have heatmaps does not affect the results.

7.3 Treatment Definition

Duration Cut-Offs If the effect really does reflect longer videos being harder to watch, then results from Equation 11 should be very robust in changing the treatment definition. They are, as shown in Figure B.17. Figure 9 shows the overall changes across all videos as a function of length, following Callaway, Goodman-Bacon, and Sant'Anna (2024). There is a sharp increase in the average change in skipping from 0 to 10 minutes, and then a general gradual increase in the propensity to skip through videos. Longer videos in general are skipped through more, which is what a lowering of attention span would predict. This result additionally assuages concerns that there may be some mechanical reason that longer videos see a greater change in clicks.[68] A mechanical effect would result in a linear change as length changed, but this is not the case. Figure B.18 additionally controls for duration linearly and the results do not change.

Figure 9 shows that my choice of cut-off is not driving the fundamental pattern. Results are also robust to allowing topic × channel and topic × treated fixed effects and additional controls, as shown in Figure B.18. Figure B.19 shows results are robust to reweighing the measure of clicks by log views, or normalising by daily total number of clicks.

Endogenous Video Length One important caveat with Equation 11 is that the length of a video is not exogenous. The crucial issue would be if channels change duration thresholds such that less interesting videos are now longer driving the effect. Logically, it would be unclear why channels might do this in response to the introduction of *Shorts*, and unclear how this would generate the observed effects. Figure B.35 shows there is no change in the probability a video is marked as treated, suggesting duration does not vary.

Nonetheless, since I have transcript data for the videos in my sample I run the usual double-debiased machine learning approach following Chernozhukov et al. (2018). It not only controls for video content, but ensures that there is no change to videos I label as treated after the policy. I use random forests as my ML method of choice, primarily to avoid issues selecting correlated predictors for text data. As previously mentioned the coefficient in Table 6 is almost identical, ruling this explanation out.[69]

[68] There are only 100 slices in both long and short videos, and so slices change length with video length.

[69] As a purely descriptive complement, Figure B.33 in Appendix B.6 reports permutation importance

8 Discussion

I next examine what this increase in skipping might mean for news and media consumption more generally. I proceed by asking why the loss of context should be expected to matter for downstream outcomes, how the effect is distributed, and whether the mechanism extends beyond news.

8.1 Outcomes for News Videos

Short-form content shifts the consumption habits of news media. This trade-off between content and the overall context of a video is likely driven by greater impatience or a worse attention span, but regardless, news videos are now substantially more skipped through than before.

Why should this loss of context matter? News videos are not only a sequence of independent claims or headlines. The context of a video is precisely where qualifications or justifications live. Even heavily biased news media which is completely devoid of caveats or alternative perspectives still attempts some chain of reasoning in qualifying statements. A content-seeking viewer misses a large portion of this information.[70] Under the earlier model, users trade context for content in news media. This action converts news consumption into something closer to headline consumption, even within videos whose content has not changed at all.

I can only speculate on the downstream outcomes, but the existing literature suggests that relatively small changes to media diets can have large effects. My results show that the composition of consumption changes sharply even holding the supply of news fixed.

This matters for two reasons. First, opinions respond to how information is framed and ordered, not just to its factual content (Chong and Druckman 2007). Two viewers extracting different slices of the same video are exposed to different framings of the same story. Second, the qualifying material being skipped is not inert. Information provision experiments show that supplying fuller context measurably shifts beliefs and policy attitudes (Alesina et al. 2023), and much of what distinguishes slanted from unslanted coverage is precisely the selection of which facts are presented (Gentzkow and Jesse M Shapiro 2010). Content-seeking viewers therefore impose a form of slant on themselves. The channel supplies a story,[71] but users consume a selection of salient fragments. News consumption shifts votes and beliefs (DellaVigna and Kaplan (2007);

from a plain random forest fit directly on transcript words, showing which words are most associated with skipping across the full sample.

[70] Restricting to videos where at least 80% of sentences cross slice thresholds, shown in Figure B.36, confirms that this skipping severs arguments mid-sentence rather than at natural breaks.

[71] Of course, many of these channels will be heavily biased to begin with. This could soften the outcome if there was little meaningful context to begin with. Alternatively it may worsen it if all claims are more likely to be taken at face value as they are no longer judged as heavily by the quality of their argument.

Martin and Yurukoglu (2017)), and so such a change to attention could have impacts on political outcomes.

If beliefs are increasingly formed from decontextualised fragments, we could expect them to be more extreme, less accurate, or held with unwarranted confidence. However, measuring these beliefs directly is beyond the scope of the heatmap data, and is the natural next step in this research agenda. Nonetheless, for policy makers already concerned with the "attention economy" (Wu 2018), the results imply that short-form content degrades the context stock consumed within existing news videos. Standard measures of news consumption (views, subscriptions, or time spent) would entirely miss this margin.

8.2 Asymmetries in the Effect

Both left and right wing channels[72] are affected, but the increase in skipping is larger for right wing channels and podcasts, as shown in Figure 18. Users more susceptible to *Shorts* will consume news media differently. If this susceptibility varies between groups, then *Shorts* creates a wedge in the context consumed within news media. Some groups form beliefs from increasingly decontextualised news while others do not.

I also compare the increase in skipping between alternative and traditional media. It is present and quite similar in both, as shown in Figure 19. However, a key feature of traditional media on *YouTube* is the short news segment: most videos posted by Fox News, for example, are below 10 minutes. Alternative media is usually longer, less scripted, less edited and more differentiated. Since the effect is sustained precisely for long, less-edited content, the rise of short-form content drives a further wedge between those who consume traditional media on the platform[73] and those who consume alternative media. To the extent that policy makers care about which parts of the news ecosystem are retaining viewers' attention, this compositional shift is itself an outcome of the policy.

8.3 Individual Welfare

Even if users get more welfare overall from *Shorts*, news consumption has less context. It is not clear whether welfare is higher however. Consider that consumption is almost certainly naive. A user could instead consume *Shorts after* a fixed amount of usual videos, and this would result in a welfare improvement according to the model. By swapping *Shorts* consumption to the end of a session, users then have the choice to either skip through anyway or potentially watch long videos as usual.[74] That this does not happen, alongside results from Bao (2025) which suggest addiction, implies that users do not seem to correctly optimise when consuming short-form content.

[72] As discretely defined; it is not vital for these results.

[73] They are of course very different to begin with.

[74] An alternate sequence whereby on some browsing sessions users only consume *Shorts* and on others users only consume normal videos would have a similar effect.

Empirical evidence suggests there is a clear welfare loss relative to the first best under *Shorts* where consumption of long videos happens strictly before *Shorts*.[75] Whether this naive consumption is enough to lower welfare relative to a regime with no short-form content at all is unclear and untestable in this setting. Regardless, users uniquely fail to optimally consume *Shorts* and consume less context in news videos as a result.

A potential solution could be increasing switching costs between long and short videos. If users do not internalise the context cost of consuming a short they might consume a *Short*, then a long video in alternation, lowering the value of context. If instead short form content was fully separate with no cross linking to normal videos then users may find it easier to more efficiently sequence their consumption. This is examined formally in Appendix C.4 but could be very difficult to incetivise in practice.

8.4 Effects on Entertainment

Whilst news videos are important, the analysis suggests a behavioural mechanism that might affect other types of content. To examine this further, I take data from *MrBeast*, the largest *YouTube* channel in the world.[76] As we can see in Figure B.37, there is suggestive evidence that the effect on skipping is additionally present for his videos. Interestingly, we see a linear decay of the effect over time. This is unsurprising as *MrBeast* is famous for attempting to optimise attention to his videos (Forbes 2024) and this could reflect a re-adjustment to more attention grabbing content not pursued by news media. It could also be a generalised trend, as this is purely descriptive.

Heatmaps on *MrBeast's* channel suggest exposure to short-form content has wide-reaching consequences above and beyond news consumption. Furthermore, this is additional evidence ruling out substitution due to limited time. A user who highly values *Shorts* may still make an effort to watch the news. It seems less reasonable for time-constrained users to make sure they watch a *MrBeast* video but skip through instead. Rather, users want to watch a *MrBeast* video but are fundamentally affected by exposure to *Shorts*, exactly the attention externality this paper documents throughout the news setting.

9 Conclusion

The way online news is consumed on *YouTube* has fundamentally shifted since 2021, due to the introduction of short-form content. "Brainrot" seems to be the cause. Short-form content lowers users' attention spans and increases skipping. Users directly substitute

[75] This assumes that there is no long term damage from short-form content exposure, but given the immediate stable effect there does not seem to be at least for news consumption.

[76] He has 500 million subscribers and 133 billion total views as of July 2026.

their attention from shorter videos to short-form news content when it becomes available, but cannot do the same for long videos, leading to sustained effects in equilibrium. Mechanism tests suggest content seeking behaviour, where users avoid the longest, densest videos after exposure. A theoretical model alongside structural estimates suggest viewers value context at least 28% less relative to content in response. The effect is robust across a wide range of alternative comparison groups and specifications, and is clearly visible in the raw data.

The model suggests these results reflect more than a passive taste shift. Naive users over-consume *Shorts* as they do not internalise the cost that exposure imposes on their own capacity for context formation. This generates a genuine welfare loss rather than a simple change in preferences at least for their consumption of news videos. Original concern from journalists (Guardian 2024; Times 2024) implied that short-form content might lead to irreversible damage to an entire generation. However, if instead this reflects repeated short-term effects sustained through continued exposure, better blocking options at the user level could suffice to reverse the change in skipping. Even so, users choose to watch *Shorts*, and if they do internalise the context loss they may care about consuming context in news videos less than policy-makers do. Encouraging blocking may be a complex task which I leave for future research.

Short-form content does not just compete for our limited attention, but additionally degrades our capacity to sustain it. This has fundamentally changed the way online news is consumed.

References

Acemoglu, D., & Finkelstein, A. (2008). Input and technology choices in regulated industries: Evidence from the health care sector. *Journal of Political Economy*, *116*(5), 837–880.

Alesina, A., Miano, A., & Stantcheva, S. (2023). Immigration and redistribution. *The Review of Economic Studies*, *90*(1), 1–39.

Allcott, H., Gentzkow, M., & Song, L. (2022). Digital addiction. *American Economic Review*, *112*(7), 2424–2463.

Aridor, G. (2025). Measuring substitution patterns in the attention economy: An experimental approach. *The RAND Journal of Economics*, *56*(3), 302–324.

Aridor, G., Jiménez-Durán, R., Levy, R., & Song, L. (2024). The economics of social media. *Journal of Economic Literature*, *62*(4), 1422–1474.

Ash, E., Galletta, S., Masera, F., & Zhang, L. (2026). *The joe rogan effect: Politicized podcasts and the youth gender voting gap* [Working paper].

Ash, E., & Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, *15*, 659–688.

Bao, R. (2025). "just one more clip": Short videos, big self-control problems.

Beknazar-Yuzbashev, G., Jiménez-Durán, R., McCrosky, J., & Stalinski, M. (2025). *Toxic content and user engagement on social media: Evidence from a field experiment* (Working Paper No. 11644). CESifo.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Bia, M., Huber, M., & Lafférs, L. (2024). Double machine learning for sample selection models. *Journal of Business & Economic Statistics*, *42*(3), 958–969.

Billingsley, P. (2012). *Probability and measure: Anniversary edition* (Vol. 446). wiley Hoboken, NJ.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, *91*(6), 3253–3285.

Brachet, T., David, G., & Drechsler, A. M. (2012). The effect of shift structure on performance. *American Economic Journal: Applied Economics*, *4*(2), 219–246.

Braghieri, L., Levy, R., & Makarin, A. (2022). Social media and mental health. *American Economic Review*, *112*(11), 3660–3693.

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024). *Difference-in-differences with a continuous treatment* (tech. rep.). National Bureau of Economic Research.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, *225*(2), 200–230.

Chen, H., & Suen, W. (2023). Competition for attention and news quality. *American Economic Journal: Microeconomics*, *15*(3), 1–32.

Chen, Y., Li, M., Guo, F., & Wang, X. (2023). The effect of short-form video addiction on users' attention. *Behaviour & Information Technology*, *42*(16), 2893–2910.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, *99*(4), 1145–1177.

Chong, D., & Druckman, J. N. (2007). Framing theory. *Annu. Rev. Polit. Sci.*, *10*(1), 103–126.

Coviello, D., Ichino, A., & Persico, N. (2014). Time allocation and task juggling. *American Economic Review*, *104*(2), 609–623.

DellaVigna, S., & Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, *122*(3), 1187–1234.

DellaVigna, S., & La Ferrara, E. (2015). Economic and social impacts of the media. In S. P. Anderson, J. Waldfogel, & D. Strömberg (Eds.), *Handbook of media economics* (pp. 723–768, Vol. 1). Elsevier.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *Journal of Finance*, *64*(2), 709–749.

Durante, R., Pinotti, P., & Tesei, A. (2019). The political legacy of entertainment tv. *American Economic Review*, *109*(7), 2497–2530.

Enikolopov, R., Makarin, A., & Petrova, M. (2020). Social media and protest participation: Evidence from Russia. *Econometrica*, *88*(4), 1479–1514.

Forbes. (2024). 8 data-driven secrets behind mr beast's youtube empire. *Forbes*. https://www.forbes.com/sites/ianshepherd/2024/10/18/the-8-data-driven-secrets-behind-mrbeasts-youtube-empire/

Fortune. (2026). Despite his $2.6 billion net worth, mrbeast says he's having to borrow cash and doesn't even have enough money in his bank account to buy mcdonald's. *Fortune*. https://fortune.com/2026/01/13/mrbeast-jimmy-donaldson-billionaire-net-worth-in-the-red-has-to-borrow-cash-money/

Freimane, M. (2024). Gender bias, feedback, and productivity. *Job Market Paper*.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, *129*(4), 1661–1710.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574.

Gentzkow, M., & Shapiro, J. M. [Jesse M]. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, *78*(1), 35–71.

Gentzkow, M., Shapiro, J. M. [Jesse M.], & Sinkinson, M. (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review*, *101*(7), 2980–3018.

Gibson, M., & Shrader, J. (2018). Time use and labor productivity: The returns to sleep. *Review of Economics and Statistics*, *100*(5), 783–798.

Guardian, T. (2024). Is doom scrolling really rotting our brains? the evidence is getting harder to ignore. *The Guardian*. https://www.theguardian.com/commentisfree/2024/dec/09/brain-rot-word-of-the-year-reality-internet-cognitive-function

Gylfason, G. (2026). *Doomscrolling: Tiktok and news consumption* [Working paper].

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1–3), 337–386.

Kelly, C. A., & Sharot, T. (2025). Web-browsing patterns reflect and shape mood and mental health. *Nature Human Behaviour*, *9*(1), 133–146.

Lacetera, N., Pope, D. G., & Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, *102*(5), 2206–2236.

Lin, B.-H., Chung, Y.-J., Cheng, H.-Y., Yen, Y.-T., Li, C.-C., & Cherng, F.-Y. (2024). Understanding the effects of short-form videos on sustained attention. *Extended abstracts of the CHI conference on human factors in computing systems*, 1–8.

Loewenstein, G., & Wojtowicz, Z. (2025). The economics of attention. *Journal of Economic Literature*, *63*(3), 1038–1089.

Ludwig, J., Mullainathan, S., & Rambachan, A. (2025). *Large language models: An applied econometric framework* (tech. rep.). National Bureau of Economic Research.

Maćkowiak, B., & Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, *99*(3), 769–803.

Mahmoudi, M., Taghiyareh, F., & Arian, K. (2024). Effect of short-form videos on student attention. *2024 11th international symposium on telecommunications (IST)*, 774–779.

Martin, G. J., & Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, *107*(9), 2565–2599.

Matejka, F., & McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, *105*(1), 272–298.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, *69*(2), 307–342.

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*, 173–182.

Phipps, G., & Honeycutt, J. (2025). Does increased use of short-form video content increase attention problems? *Evidence-Based Practice*, *28*(11), 15–16.

Puaponpong, W., Thongdee, N., & Chutanopmanee, V. (2023). Correlation between the use of short-form video and everyday life attention in thai high school students. *International Journal of Current Science Research and Review*, *6*(8), 5983–5991.

Shanavas, R., Faizan, M., Singh, M., Naqvi, F., Samudrala, K., Chhawal, J., Ahmed, T., Janabi, H., Ramesh, N., Balaji, S. N., et al. (2025). The effects of short-form media consumption on attention and focus: A pilot study. *Innovations in Digital Health, Diagnostics, and Biomarkers*, *5*(2025), 44–48.

Sievertsen, H. H., Gino, F., & Piovesan, M. (2016). Cognitive fatigue influences students' performance on standardized tests. *Proceedings of the National Academy of Sciences*, *113*(10), 2621–2624.

Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy*, *49*, 317–356.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690.

Stango, V., & Zinman, J. (2014). Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. *Review of Financial Studies*, *27*(4), 990–1030.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, *225*(2), 175–199.

Times, T. N. Y. (2024). If you know what brainrot means, you might already have it. *The New York Times*. https://www.nytimes.com/2024/06/13/style/brainrot-internet-addiction-social-media-tiktok.html

Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, *86*(3), 859–889.

Train, K. E. (2009). *Discrete choice methods with simulation* (2nd). Cambridge University Press.

Violot, C., Elmas, T., Bilogrevic, I., & Humbert, M. (2024). Shorts vs. regular videos on youtube: A comparative analysis of user engagement and content creation trends. *Proceedings of the 16th ACM web science conference*, 213–223.

Wei, M., Liu, J., Wang, H., Li, Q., & Dong, G.-H. (2026). Fragmented learning from short videos modulates neural activity and connectivity during memory retrieval. *npj Science of Learning*.

Wojtowicz, Z., Chater, N., & Loewenstein, G. (2019). Boredom and flow: An opportunity cost theory of motivational attention. *Available at SSRN 3339123*.

Wu, T. (2018). Blind spot: The attention economy and the law. *Antitrust LJ*, *82*, 771.

Zhong, Y. (2024). Working under distractions.

10 Tables

Table 1: Channel Video Summary

Channel	Videos	Non-Short Videos w/ Heatmap	Slices
Ben Shapiro	8,607	4,013	401,300
Brian Tyler Cohen	5,836	4,412	441,200
CNN	8,407	3,000	300,000
Candace Owens	1,553	328	32,800
Charlie Kirk	1,885	772	77,200
Conservative Twins	2,228	1,577	157,700
DailyWire+	8,200	3,825	382,500
David Pakman Show	10,475	2,377	237,700
Destiny	3,420	2,627	262,700
Fox News	8,983	4,381	438,100
HasanAbi	2,899	2,296	229,600
Jordan B Peterson	1,560	1,066	106,600
Lex Fridman	805	579	57,900
MSNBC	9,520	1,093	109,300
Philip DeFranco	2,273	1,698	169,800
PowerfulJRE	3,260	2,953	295,300
Sam Harris	461	265	26,500
StevenCrowder	2,190	1,722	172,200
The Majority Report	9,398	2,071	207,100
The Rubin Report	9,159	4,801	480,100
Theo Von	1,234	604	60,400
Tim Pool	3,734	1,662	166,200
Timcast	9,056	7,276	727,600
Tucker Carlson	1,472	337	33,700
Turning Point USA	7,226	491	49,100
Total	117,842	56,226	5,622,600
AajTak	18,461	3,843	384,300
ABP NEWS	9,596	818	81,800
Dhruv Rathee	778	766	76,600
IndiaTV	9,242	1,487	148,700
Sandeep Maheshwari	639	638	63,800
MrBeast	466	464	46,400

Notes: Table 1 shows the list of channels in my sample. I aim to find a mix of traditional and alternative, left and right, podcasts and daily shows. Some channels have a mix of longer and shorter videos, such as Ben Shapiro or CNN, whereas others such as Joe Rogan (PowerfulJRE) mostly release longer content. The number of videos for each channel in my sample is shown in the first column. For CNN, MSNBC, Fox News, The Majority Report and The Rubin Report this is a subsample of 1 in N videos to get close to 10,000. The second column displays the videos for which I have heatmap data and are not short-form content. This drops off both due to some videos not having enough views to generate a heatmap, and some videos with heatmaps being short-form content. The third column reports the number of slices (the base N in my regressions, before duration restrictions or slice restrictions). The Indian channels (AajTak, ABP NEWS, Dhruv Rathee, IndiaTV and Sandeep Maheshwari) are only included when estimating Equation 10, hence are displayed separately. MrBeast is displayed separately as an additional benchmark channel.

Table 2: Descriptive Statistics

Variable	Mean	SD	Median	Min	Max
Activity at Slice	0.148	0.178	0.0779	0.00	1
Video Duration (Seconds)	1,101	2,152	359	4	42,840
Views	489,307	1,315,063	159,354	15.00	65,074,186
Greater than 1 hour	0.082	0.275	0	0	1
Likes	18,043	44,551	5,868	0	2,307,836
Transcript Available	0.864	0.343	1	0	1

Notes: Table 2 shows simple descriptive statistics for the 5,622,600 slices in my sample. For video measures, i.e. all objects that are not activity, I take the average at the video level to ensure the standard deviation correctly reflects variation.

Table 3: Effect of Shorts on Skipping: Before and After

	Proportion Clicking					
	All Videos			*Videos > 30 min*		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Effect	0.0258***	0.0396***	0.0396***	0.0679***	0.0880***	0.0888***
	(0.0058)	(0.0059)	(0.0060)	(0.0147)	(0.0154)	(0.0167)
Uploader FEs	Yes	Yes	No	Yes	Yes	No
Uploader × Topic FEs	No	No	Yes	No	No	Yes
Flexible Controls	No	Yes	Yes	No	Yes	Yes
Observations	5,622,600	5,622,600	5,096,100	1,422,000	1,422,000	1,371,800
Dependent var. mean	0.148	0.148	0.141	0.148	0.148	0.141

Notes: Table 3 reports the overall effect of *Shorts* exposure, with treatment in this setting defined by the uploader's month of first *Shorts* exposure. Estimation follows Sun and Abraham 2021, comparing a channel exposed to *Shorts* at N months since first video to one not exposed to *Shorts* at N months since first video. This is broadly a before and after comparison, and does not rule out contemperaneous shocks. Columns (1) to (3) use all videos. Columns (4) to (6) restrict to videos longer than 30 minutes. Within each panel, the first column includes only uploader fixed effects. The second adds uploader-specific cubic polynomials in views and likes interacted with relative month, absorbing uploader-specific popularity trends around treatment. The third replaces uploader fixed effects with uploader × dominant-topic fixed effects, allowing the level of retention to vary flexibly across content categories within uploader. The estimated effect is stable across specifications and increases on the long-video subsample driven by the effect being sustained over time. Standard errors are clustered at the video and uploader level and shown in parentheses.

Table 4: Effect of Shorts on Skipping: Across Markets

	Proportion Clicking			
	All Videos		*Videos > 30 min*	
	(1)	(2)	(3)	(4)
Treatment Effect	0.0206**	0.0215**	0.0603***	0.0605***
	(0.0069)	(0.0068)	(0.0125)	(0.0108)
Uploader FEs	Yes	Yes	Yes	Yes
Flexible Controls	No	Yes	No	Yes
Observations	6,050,600	6,050,600	1,518,100	1,518,100
Dependent var. mean	0.147	0.147	0.156	0.156

Notes: Table 4 reports the overall effect of *Shorts* exposure on skipping across markets using both US and Indian channels, with treatment defined by the uploader's month of first *Shorts* exposure (September 2020 for Indian channels and March 2021 for US channels). Estimation follows Sun and Abraham 2021 with the caveat of assuming that treatment effect heterogeneity and dynamics in the US specifically and India do not cancel out with common shocks. Columns (1) to (2) use all videos. Columns (3) to (4) restrict to videos longer than 30 minutes. The first column in each panel includes only uploader fixed effects. The second adds uploader-specific cubic polynomials in views as in the previous table. Standard errors are clustered at the video and uploader level and shown in parentheses.

Table 5: Across Videos: Robustness to Fixed Effects

	Proportion Clicking			
	(1)	(2)	(3)	(4)
Post × Treated (Long)	0.0693***	0.0699***	0.0622***	0.0630***
	(0.0158)	(0.0141)	(0.0155)	(0.0139)
Day FEs	Yes	Yes	Yes	Yes
Uploader FEs	No	Yes	Yes	Yes
Slice FEs	No	Yes	Yes	Yes
Uploader × Treated × Slice FEs	No	No	Yes	Yes
Month × Uploader × Slice FEs	No	No	Yes	Yes
Uploader × Treated × Slice × Topic FEs	No	No	No	Yes
Month × Uploader × Slice × Topic FEs	No	No	No	Yes
Flexible Views × Slice Controls	No	No	Yes	Yes
Observations	5,622,600	5,622,600	5,622,600	5,023,500
Dependent var. mean	0.148	0.148	0.148	0.141

Notes: Table 5 compares skipping through videos greater than one hour and less than one hour, before and after shorts. More specifically, it shows the effect of video length on proportion clicking at a given video slice, read as an increase in skipping through videos. The results show that longer videos receive much more skipping than shorter videos. Long is defined as a video 1 hour or greater, but results are robust across any cut-off 20 minutes or more. Columns (1) to (4) estimate Equation 11 across an increasingly rich fixed effects structure. The coefficient on Post × Treated is stable and remains highly significant throughout, moving only marginally as day, uploader, slice, and topic fixed effects and flexible view controls are added. The key fixed effect of note is any Uploader × Month which partials out overall effects and identifies only the differential within channel between long and not long videos. Standard errors are clustered at the video and uploader level, and shown in parentheses.

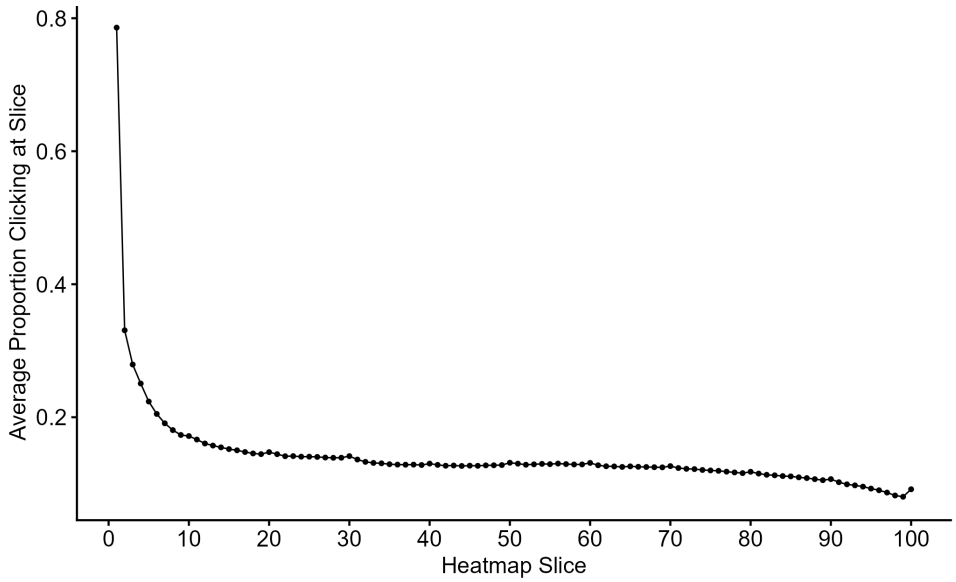
Table 6: Video-Level Robustness To Transcript Content

	Proportion Clicking at Slice (Video Level)				
	Baseline	+ Topics	DDML	DDML-Words	DDML-SSM
Post × Treated (Long)	0.0629***	0.0616***	0.0621***	0.0602***	0.0552***
	(0.0168)	(0.0159)	(0.004)	(0.004)	(0.008)
Treated × Uploader FEs	Yes	Yes	Yes	Yes	Yes
Month × Uploader FEs	Yes	Yes	Yes	Yes	Yes
Day FEs	Yes	Yes	Yes	Yes	Yes
Uploader × Dominant Topic FEs	No	Yes	N/A	N/A	N/A
Observations (Videos)	56,226	50,235	49,139	49,139	49,139
Dependent var. mean	0.148	0.141	0.141	0.141	0.141

Notes: Table 6 estimates Equation 11 at the video level. Baseline and + Topics repeat the estimation previously run at the slice level as a benchmark. Here we find the significant average effect on skipping to a slice as before. Video level data is used for ease of computation, and since this is the level at which the transcript varies. The DDML column uses R's DoubleML package on de-meaned data to estimate nuisance functions for Post × Treated and Mean Activity. Column DDML does this on the TF-IDF matrix of the top 5000 terms, further decomposed into 300 columns through SVD. DDML-Words runs the same estimation directly on the top 3000 terms as raw words. This ensures that results are not driven by changes in video content, either through different videos selecting into treatment or a general change in transcript content not captured by fixed effects. As we can see, the estimate is very similar with an even sharper confidence bound, suggesting the effect is not driven by changes in video content. Finally, DDML-SSM uses R's DoubleML package to estimate a flexible sample selection model where text data affects both the outcome and the probability of being selected. Again, the estimate does not change substantially suggesting that these results are not driven by different videos being treated or different videos being observed. As a purely descriptive complement, a plain random forest with permutation importance (no fixed effects, no causal interpretation) is separately fit on the same tf-idf word matrix to predict the proportion clicking at slice directly, identifying which words are most predictive of higher or lower skipping across the full sample of videos. Random Forests are used in all ML estimators due to their tendency to select uncorrelated predictors. Standard errors are clustered at the video level when possible, and in parentheses.

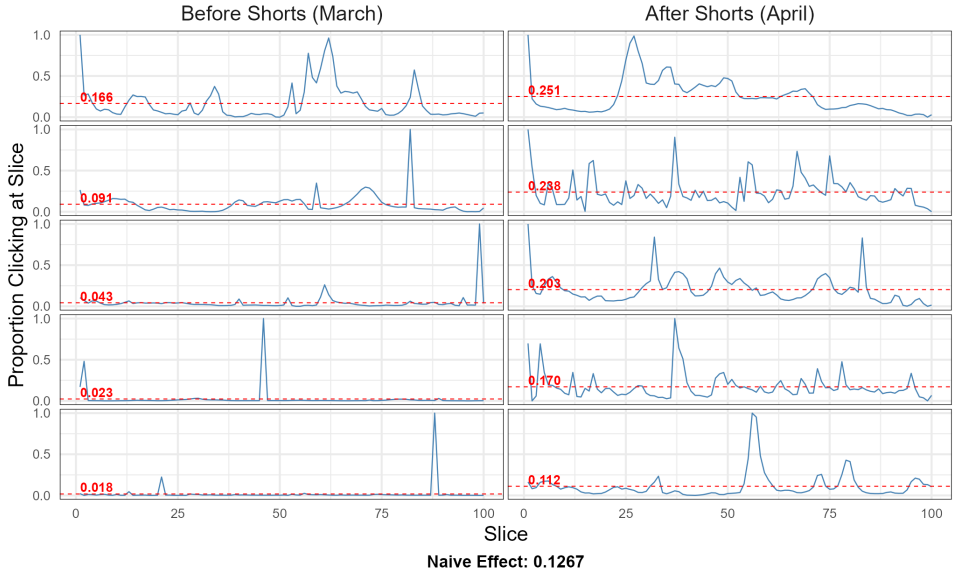
11 Figures

Figure 1: Average Heatmap Across Slices



Notes: Figure 1 shows the average heatmap across all videos i.e. the average measured clicking activity at each of the 100 slices. The heatmap is normalised to one at its most clicked section, with all other slices in relation to this one. On average, clicking activity is higher towards the start of a video than towards the end. This is likely due to users replaying the very beginning of videos.

Figure 2: Example Heatmaps Before and After the Policy



Notes: Figure 2 shows 5 randomly selected heatmaps from *Ben Shapiro*'s videos just before, and 5 just after, the introduction of *Shorts*. This gives a sense of the variation captured by the estimating equations. All videos are taken from Ben Shapiro to ensure a reasonable comparison, taking videos from the first week of March and the first week of April. The red line displays the average proportion skipping at a slice, and there is a clear increase, driven by the clear change in how videos seem to be consumed. There is very little skipping or replaying before the policy, though some videos do have peaks and valleys. However, afterwards the average skipping is higher and the shape has changed. This is purely descriptive, but the average effects I find are in line with this type of observed change.

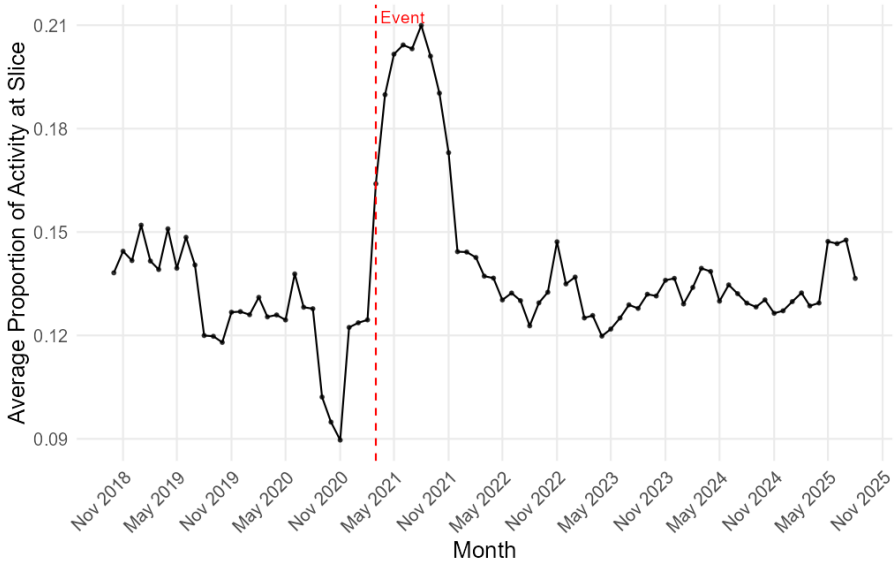
Figure 3: Measuring Skips: Sponsored Segment Heatmap



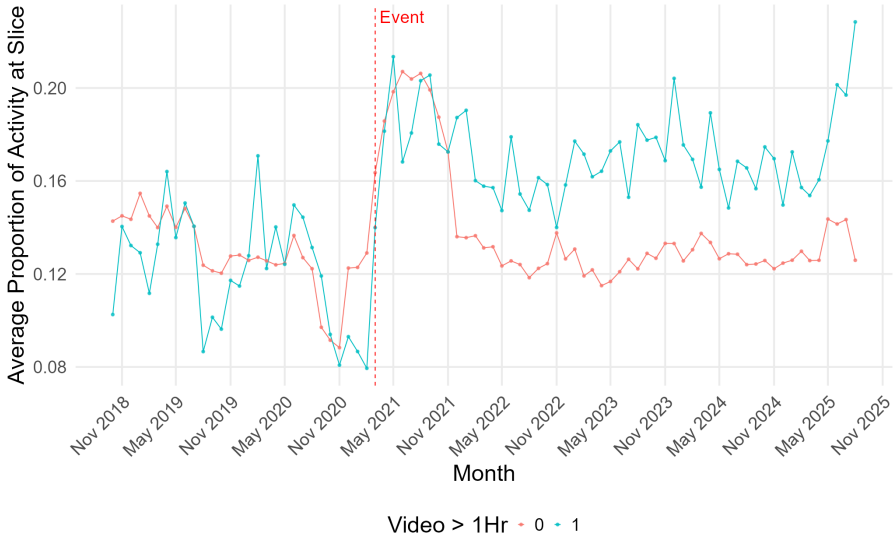
Notes: Figure 3 shows the raw attention heatmap for a video containing a sponsored segment, highlighted in yellow. This shows that the heatmap measures skipping. I do not know the exact formulation, but either the heatmap directly measures skips, or users skip past sections and then replay backwards to the start of a section. In either case, the clicks measure skipping. There is a sharp fall in activity during the sponsored segment, followed by a large mass of clicks immediately after it ends, consistent with users skipping past the sponsored content. This validates the measure, as sponsored sections are the ideal part of the video for a user to skip.

Figure 4: Heatmap Descriptives

Panel A: Mean Activity Over Time



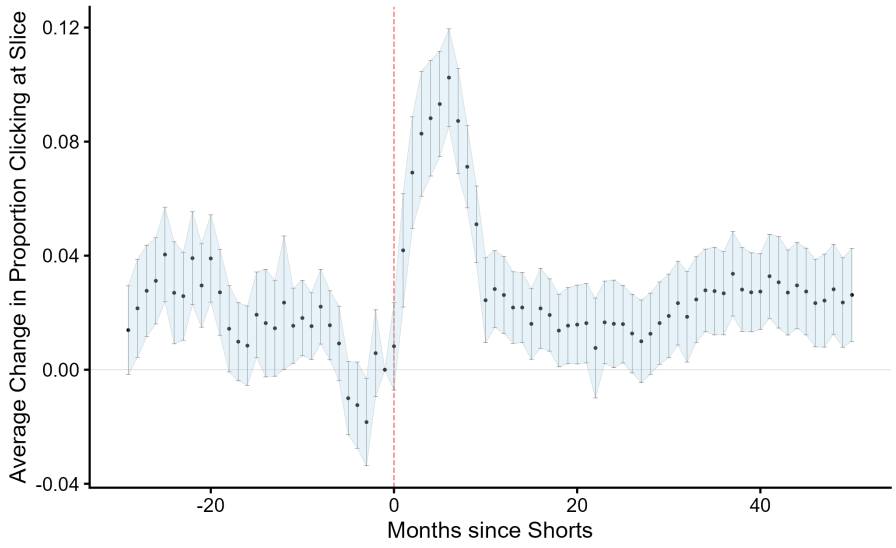
Panel B: Mean Activity Over Time, by Duration



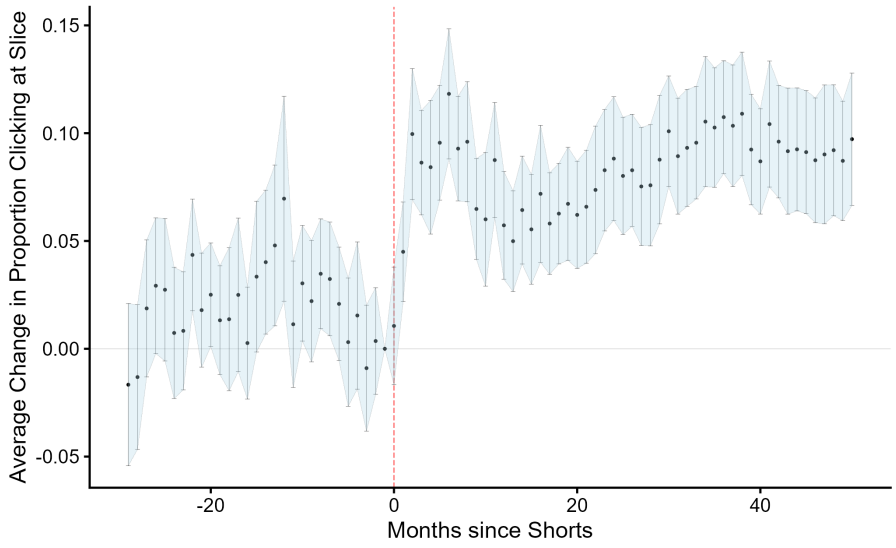
Notes: Figure 4 shows the average proportion skipping at a slice across time. Panel A is the monthly average activity across all slices in my sample. There is a clear spike at the introduction of *Shorts*. This will be sustained for longer videos, as shown in Panel B. The raw data is split by length, and we can see that for videos greater than 1 hour there is a sustained mean shift in skipping behaviour after the policy. There is clearly a shift in video watching behaviour around the introduction of *Shorts* which is sustained for longer videos. This will be further explored in Section 6.1.

Figure 5: Before and After: Event Studies

Panel A: All Videos



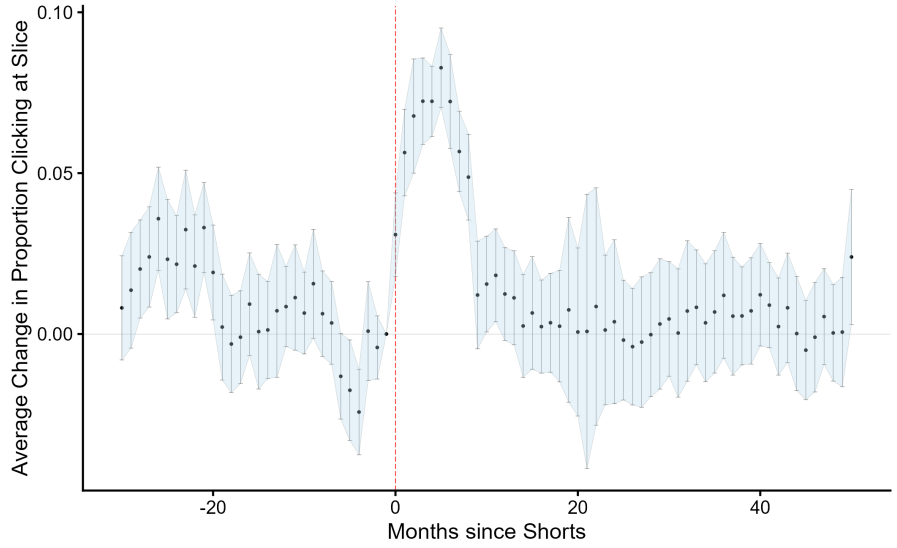
Panel B: Videos > 30 Minutes



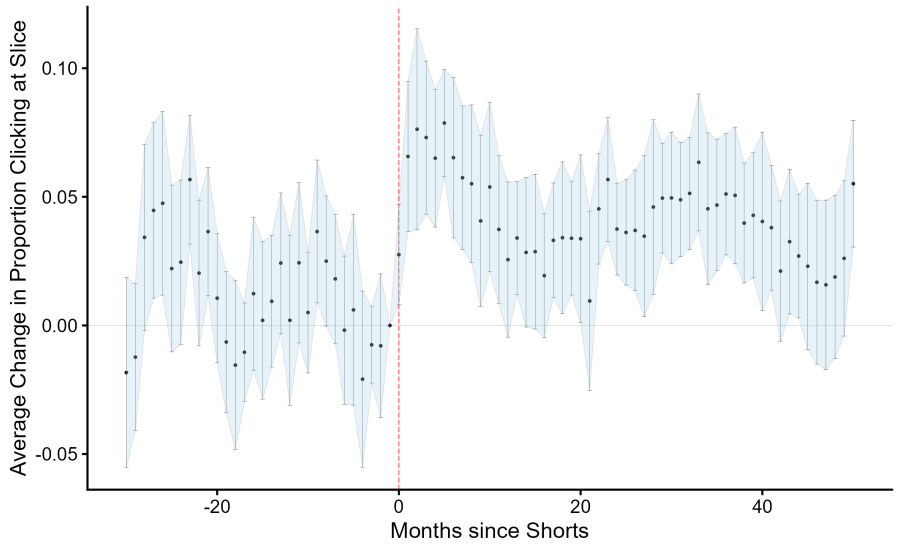
Notes: Figure 5 shows the effect of *Shorts* on the proportion skipping at a given slice. I estimate Equation 9 across all videos and videos at least 30 minutes. Estimation follows Sun and Abraham (2021). This is a simple before and after, where the control group are other channels at the same time in their life cycle. For example, this specification would compare Ben Shapiro's channel at N years old to Steven Crowder's channel at N years old, with different exposure to *Shorts*. On its own, this is descriptive. However, there seems to be a clear effect of the policy. This specification controls for slice, topic × channel and day fixed effects, and views and duration as 3rd degree polynomials. Results are robust to varying or omitting controls. Confidence intervals are clustered at the video level, at 95% coverage.

Figure 6: Attention Across Markets: Event Studies

Panel A: All Videos

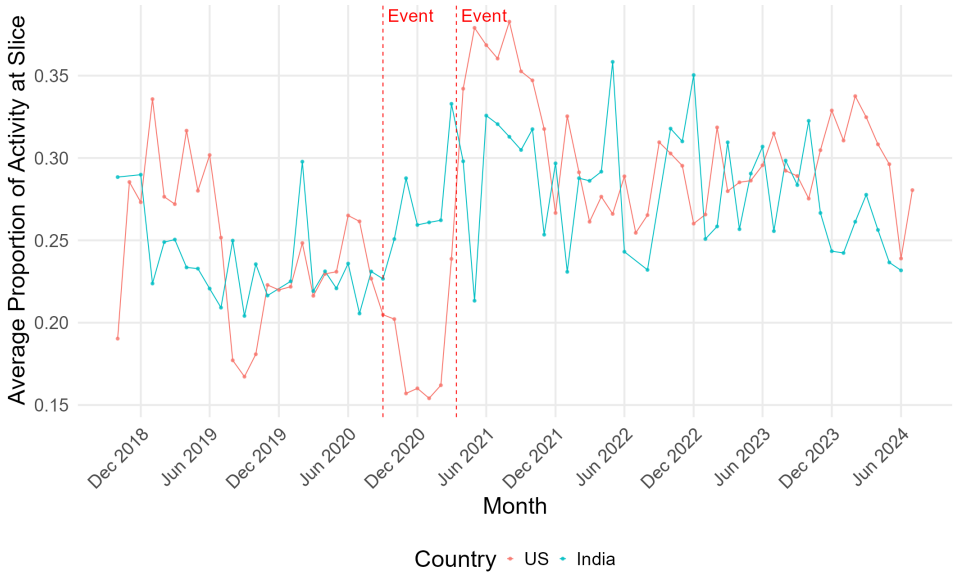


Panel B: Videos > 30 Minutes



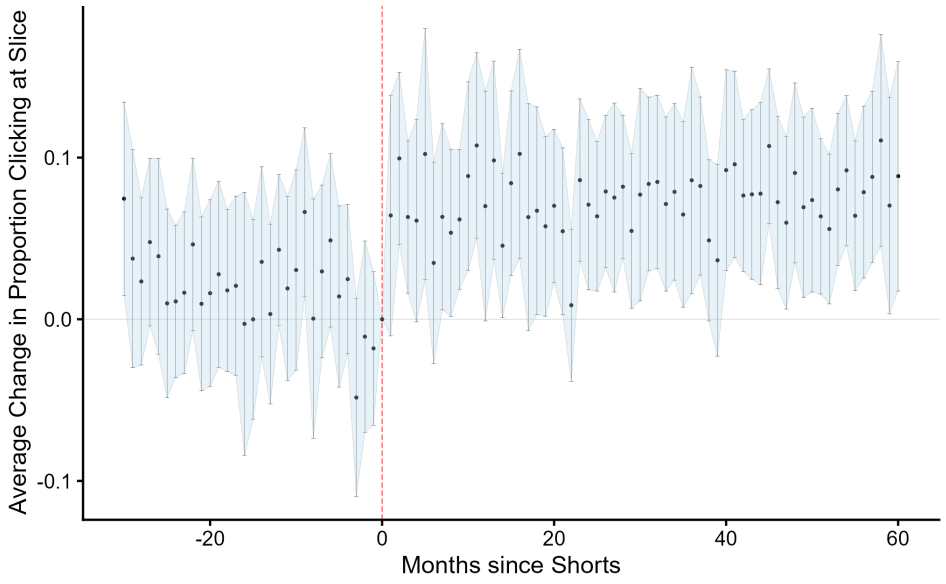
Notes: Figure 6 shows the effect of *Shorts* exposure on skipping across markets. I estimate Equation 10 across all videos and videos at least 30 minutes, comparing the staggered rollout of *Shorts*. Estimation follows Sun and Abraham 2021. There is a clear effect of the policy, ruling out some kind of platform-wide shift in how attention works or is measured, or a contemporaneous shock to attention. The results are noisier given *India* as a comparison group, but show a clear effect. The raw data shows this is driven by both the Indian and US channels, in Figure 7. Confidence intervals are clustered at the video level, at 95% coverage.

Figure 7: DiD Comparing US and India: Raw Data



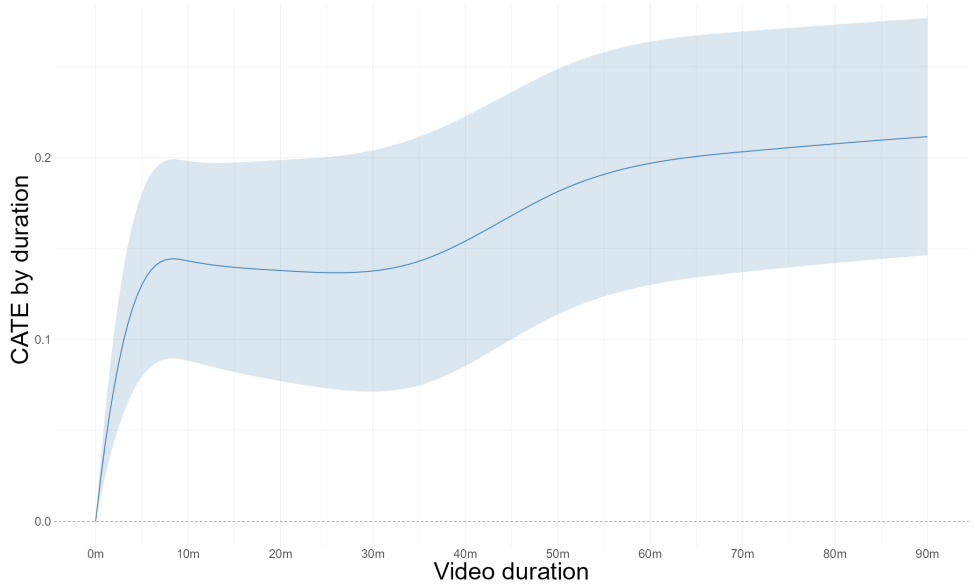
Notes: Figure 7 shows a rolling average of proportion skipping comparing US channels to the 5 Indian channels in my sample, for videos at least 20 minutes long. First, there is importantly no contemporaneous spike in March 2021 for Indian channels, ruling out measurement changes or updates driving the observed result. Second, there seems to be a less sudden but present increase in skipping when *Shorts* was released in India. *Shorts* exposure spikes skipping only in India in September 2020, and only in the US in March 2021, confirming the timing of the effect tracks the staggered rollout rather than a common platform-wide shock. This strongly suggests *Shorts* causally increased skipping. Looking at the variance of the heatmap gives also gives a strong causal pattern as shown in Figure B.14.

Figure 8: Attention Across Videos: Duration



Notes: Figure 8 compares the effect of skipping in a difference-in-differences approach for long videos to short videos. I estimate Equation 11 to both bolster causal claims and examine whether attention is the likely mechanism. Treated videos are videos greater than 60 minutes, and control are videos less than 60 minutes, and results are robust to alternative cut-offs. There is an immediate, sustained increase in skipping for longer videos *within the same channel* relative to shorter videos after the introduction of *Shorts*. This effect is consistent with an attention span effect that is more severe for content with greater contextual complexity. Longer videos are harder to focus on, and contain more contextual references. Confidence intervals are clustered at the video level, at 95% coverage.

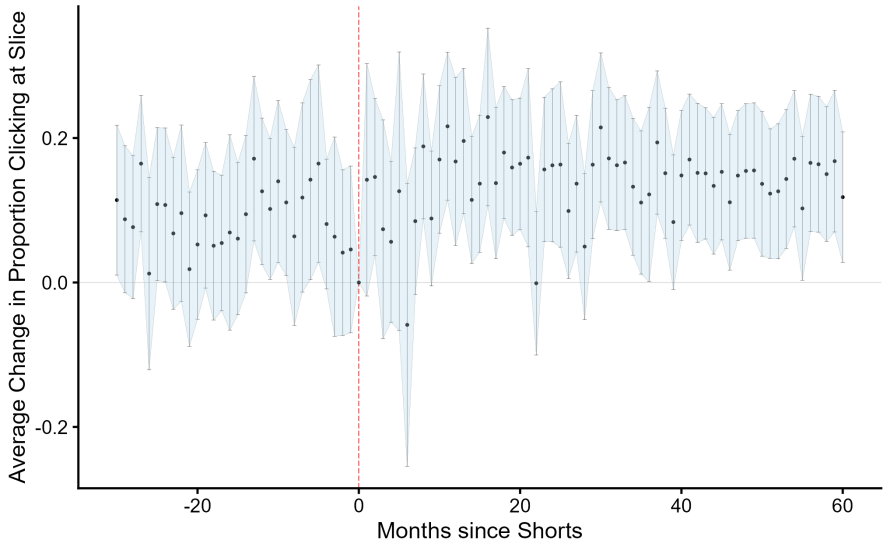
Figure 9: Attention Across Videos: Continuous Treatment by Duration



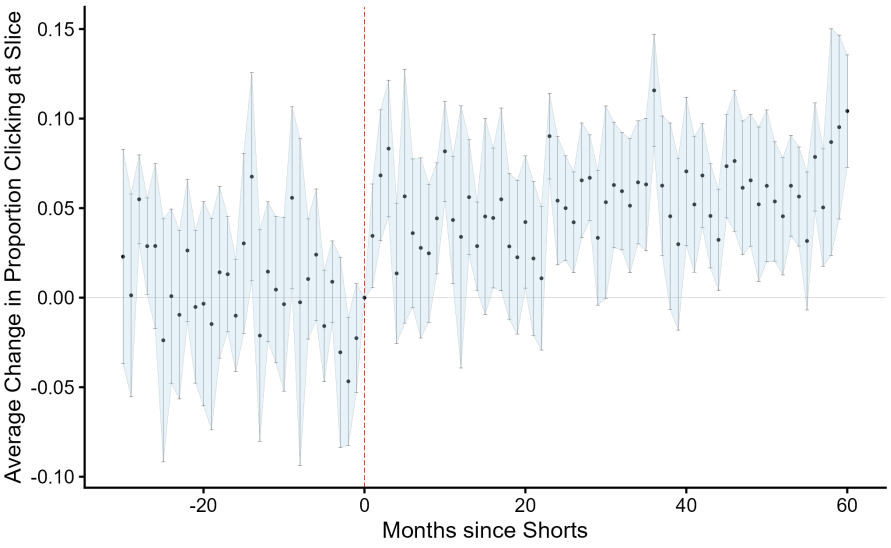
Notes: Figure 9 shows the overall change in skipping as a continuous function of video length, following Callaway, Goodman-Bacon, and Sant'Anna (2024) with B-splines. There is a sharp increase from 0 to 10 minutes, and a general gradual increase in skipping thereafter, confirming the binary treatment cut-off used elsewhere is not driving the results. Confidence intervals are clustered at the video level, at 95% coverage.

Figure 10: Duration Effect by Information Density

Panel A: WPM Below Median

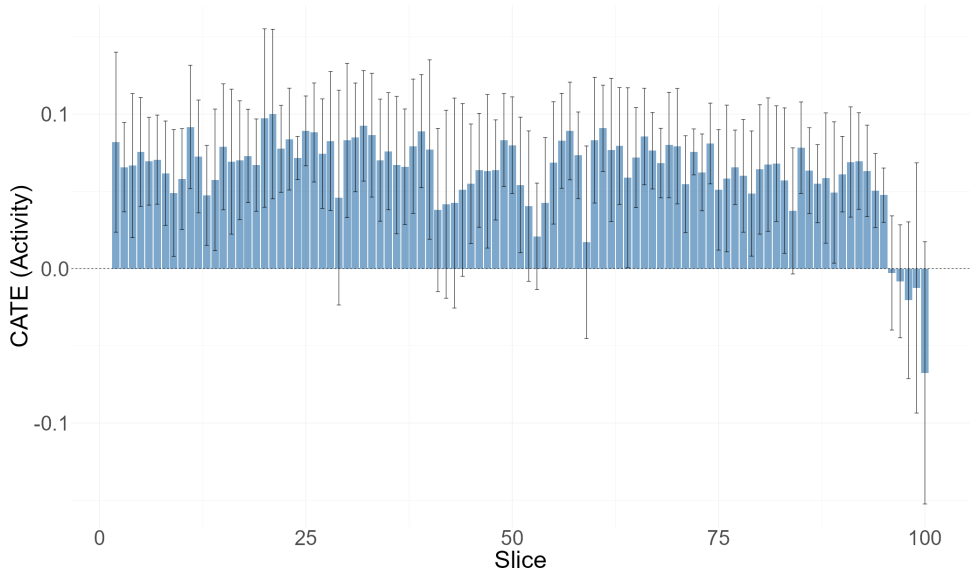


Panel B: WPM Above Median



Notes: Figure 10 estimates the effect of *Shorts* exposure on skipping, for long and short videos above and below median words per minute (WPM), a proxy for information density. The crucial comparison is whether the time differential is higher or lower for dense videos. If time is the binding constraint, viewers benefit the most from skipping through the longest, slowest videos. If attention is the binding constraint, then viewers would find the longest, most information dense videos even more beneficial to skip than just long videos. The relative effect of length is driven by skipping on the most information-dense videos, consistent with attention rather than time being the binding constraint. Confidence intervals are clustered at the video level, at 95% coverage.

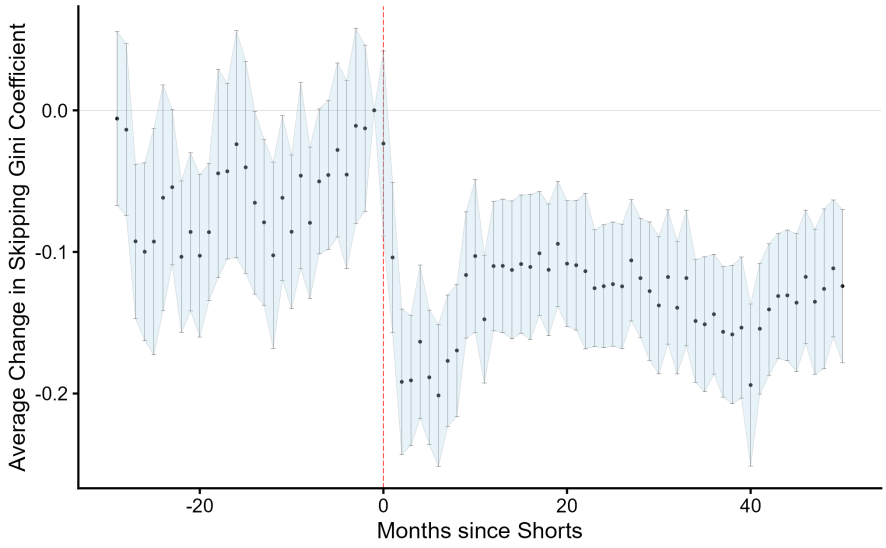
Figure 11: Effect by Slice



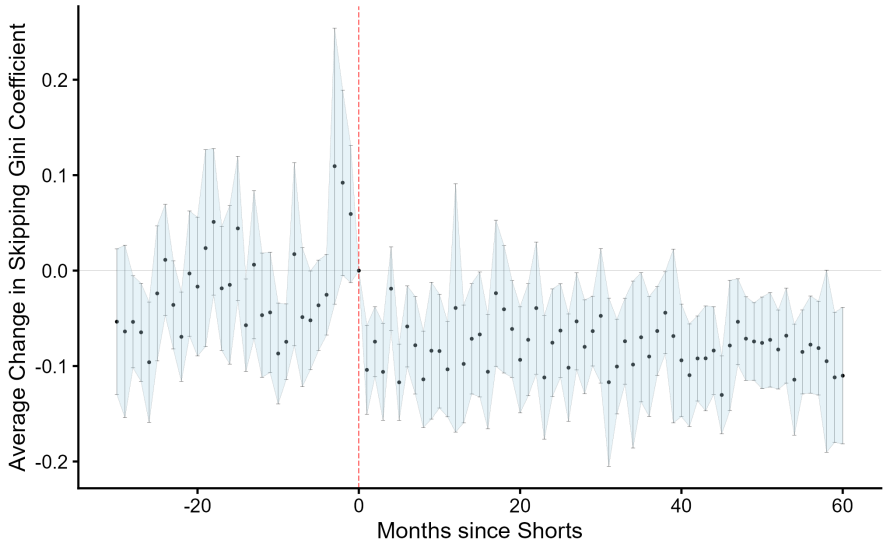
Notes: Figure 11 re-estimates Equation 9 across each individual slice from 2 to 100, for videos longer than 30 minutes. This examines how the increase in skipping is distributed within a video on average. Slice 1 is omitted as the event study reveals no increase, but the coefficient is large due to minor pre-trends and in general a pre-treatment mean of a much greater magnitude. The other event studies have all been manually checked (see Figures B.24–B.27 in the Appendix) and in almost all cases pre-trends hold strongly for both the significant and the null outcomes. As we can see, there is a consistent, similar increase across all but the last few slices. This suggests a content-seeking mechanism, where users skip throughout a video and avoid the likely low content end of a video. Confidence intervals are clustered at the uploader level, at 95% coverage.

Figure 12: Within-Video Spread: Gini Coefficient

Panel A: Overall



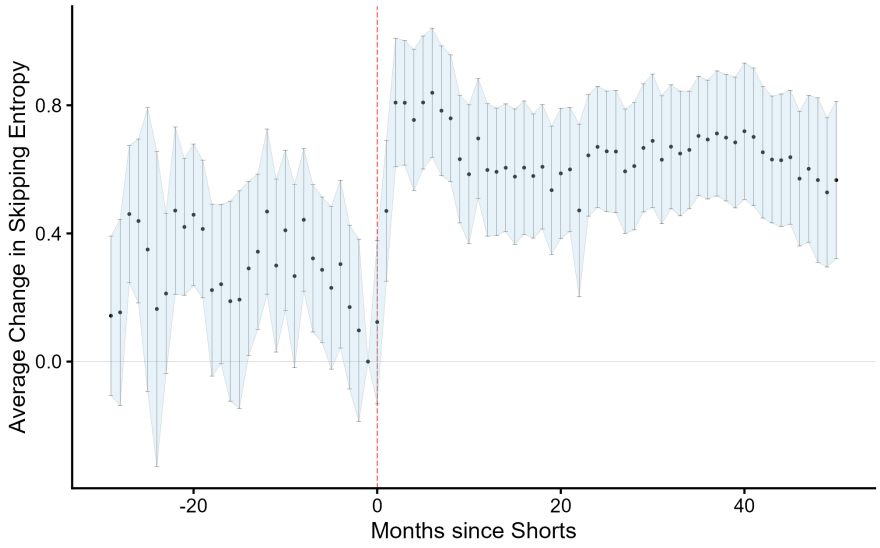
Panel B: Long Vs Short



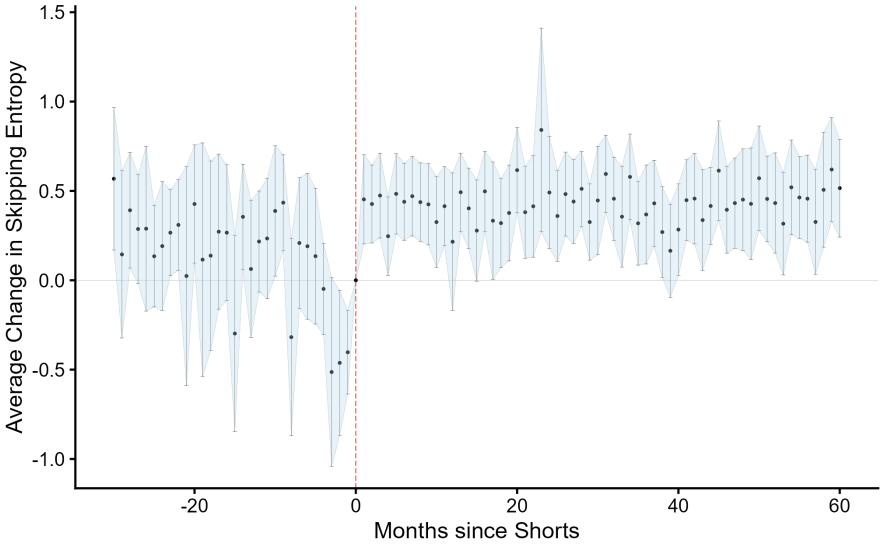
Notes: Figure 12 shows the average Gini coefficient of the video-level heatmap. Panel A shows the overall effect following Equation 9 for videos longer than 30 minutes, Panel B compares long and short videos following Equation 11. The average Gini coefficient falls after the policy, consistent with skipping becoming more uniform within a video rather than concentrated at a few cut points. This strongly rules out a selection story, as if users were splitting videos in order to watch more shorts, we should expect skipping to be mostly driven by a few large cut points. Confidence intervals are clustered at the video level, at 95% coverage.

Figure 13: Within-Video Spread: Entropy

Panel A: Overall

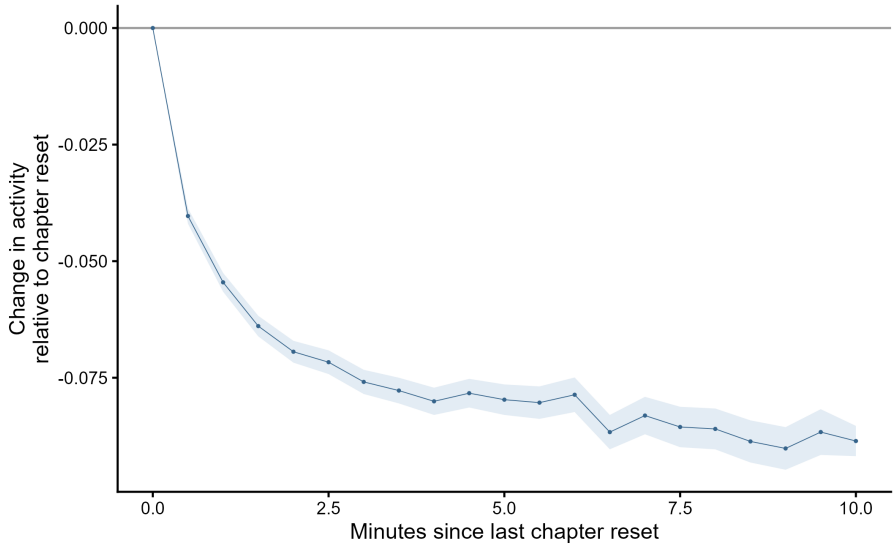


Panel B: Long Vs Short



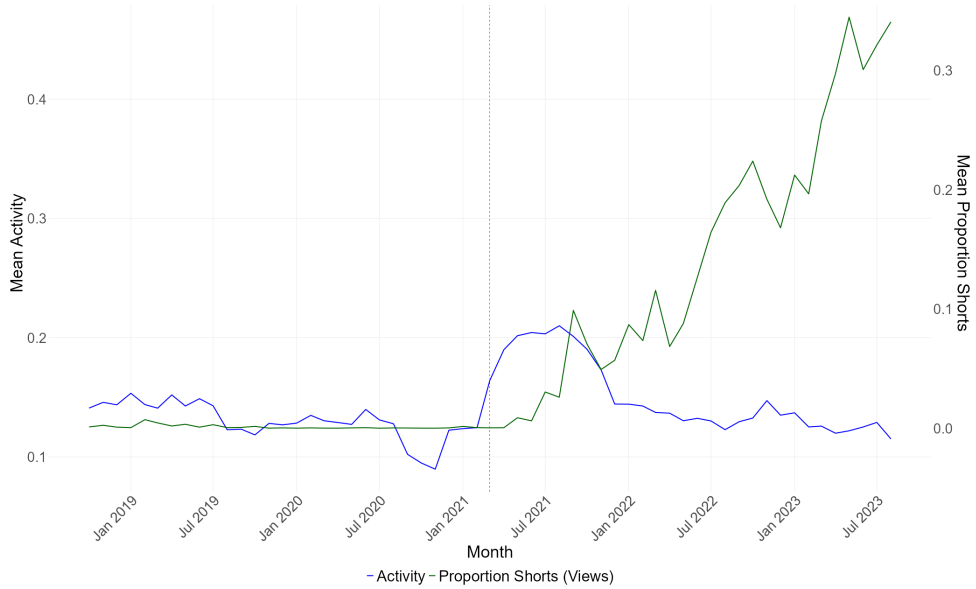
Notes: Figure 13 shows the average entropy of the video-level heatmap. Panel A shows the overall effect following Equation 9 for videos longer than 30 minutes, Panel B compares long and short videos following Equation 11. The average entropy rises after the policy, which is a measure of how uniform the heatmap is. This is again consistent with skipping becoming more uniform within a video, and corroborating the Gini coefficient result in Figure 12. Confidence intervals are clustered at the video level, at 95% coverage.

Figure 14: Change in Importance Relative to a Chapter Reset



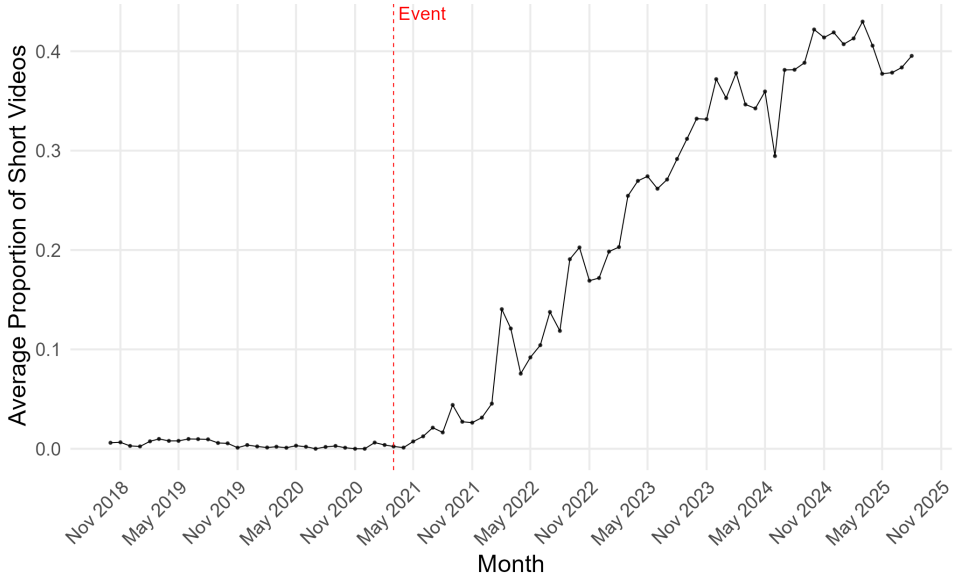
Notes: Figure 14 plots the change in activity at each depth since the last chapter reset, relative to the reset itself, before versus after *Shorts*. More specifically, this examines which slices receive relatively more skips abstracting from the overall mean shift. This is estimated by a linear probability model on chaptered videos only, with video and uploader×slice fixed effects, in 30-second depth bins out to 10 minutes plus a catch-all bin beyond that. The reset bin is the omitted reference and is fixed at 0 in both periods. We can see that there is the most skipping at a new chapter, but interestingly still more skipping after. Under some assumptions in Appendix D, this traces out the rate of context build up in online news videos. The shaded area is the 95% confidence interval from standard errors clustered at the video level.

Figure 15: Proportion of Shorts Produced by Sample Channels Over Time



Notes: Figure 15 shows the average proportion of *Shorts* in the total output of channels in my sample over time, superimposed on the average change in skipping over time for all videos. Channels do not immediately begin producing *Shorts*, with a consistent rise beginning around 7 months after the policy, which coincides with when the demand-side effect starts to decay for the shortest videos. This is suggestive evidence, but implies that the fall in skipping on the shortest videos is driven by the availability of news shorts as substitutes. Users no longer skip through the shortest videos in the sample when they have news substitutes.

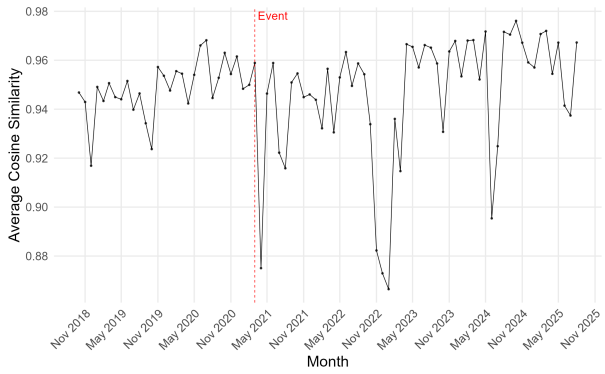
Figure 16: Average Proportion of Shorts Within Sample



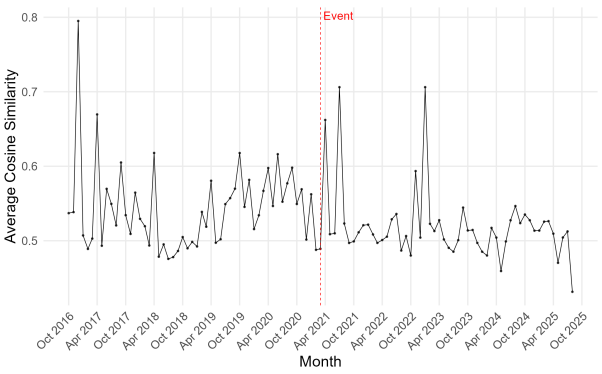
Notes: Figure 16 takes the monthly average proportion of *Shorts* relative to normal videos within channels in my sample. Before the policy, *Shorts* is very close to zero, but not identically zero, due to *YouTube*'s policy of re-tagging certain existing videos as *Shorts*. Clearly this has very little effect on the channels in my sample. After the policy there is not an immediate uptake of *Shorts*; instead there is a constant increase in the proportion of *Shorts* released by channels in my sample. If the effect were instead driven by *Shorts* bringing users with lower attention spans to the channels in my sample, we would expect to see an effect on activity that tracks the shape of this graph, as selection of low-attention individuals increases over time with a greater release of *Shorts*. We do not find this, further providing evidence towards the attention span mechanism rather than a composition story.

Figure 17: Supply Side: Text Embeddings

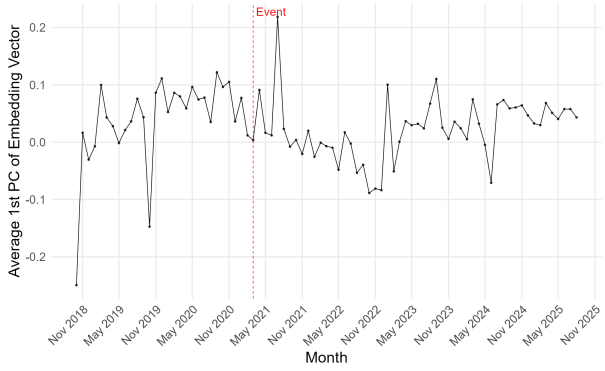
Panel A: Cosine Similarity (Across Weeks)



Panel B: Cosine Similarity (Within Weeks)



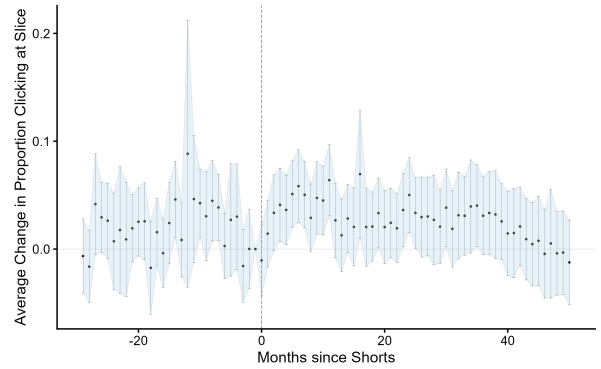
Panel C: PCA (First Principal Component)



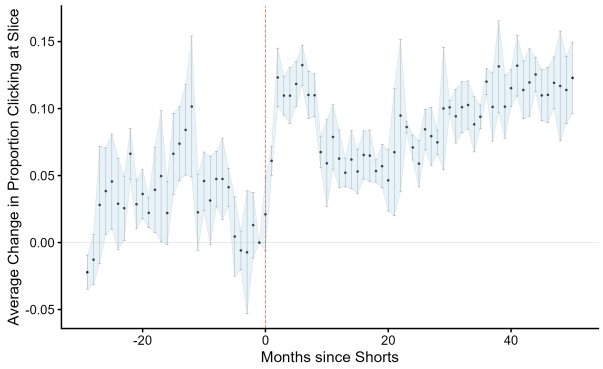
Notes: Figure 17 shows text embeddings of channels' usual videos over time. Panel A shows cosine similarity of videos across weeks, Panel B within weeks, and Panel C the first principal component of these embeddings. In all three panels there is no immediate or gradual change around the policy, indicating channels do not alter the content of their existing videos in response to *Shorts*. We see no change in the text embeddings unique to the event. The two spikes in Panel A and Panel B are not sustained, or robust to alternative centroids, and of a similar magnitude to other spikes. Therefore, it seems existing news videos did not change immediately or gradually due to *Shorts*.

Figure 18: Asymmetry: Left, Right and Podcast Channels

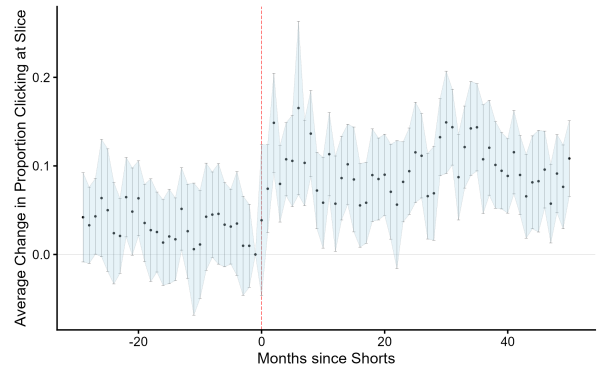
Panel A: Left



Panel B: Right



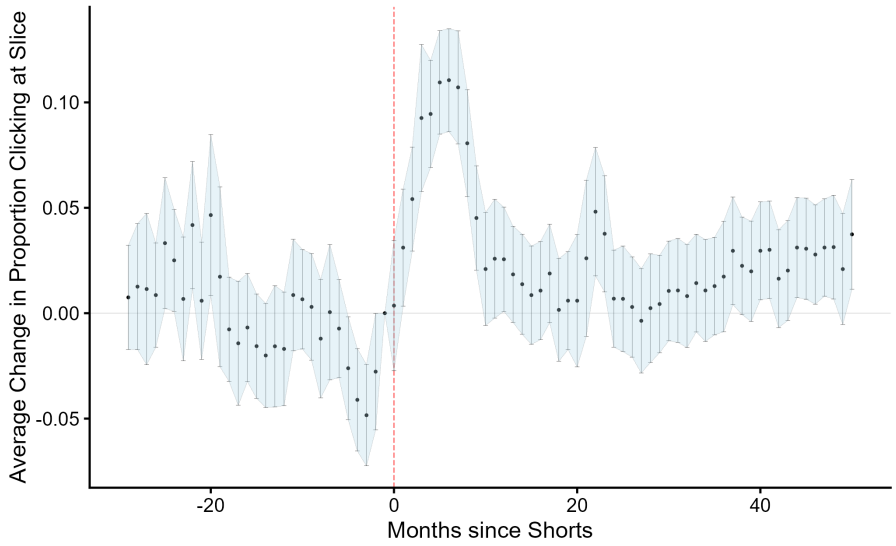
Panel C: Podcast



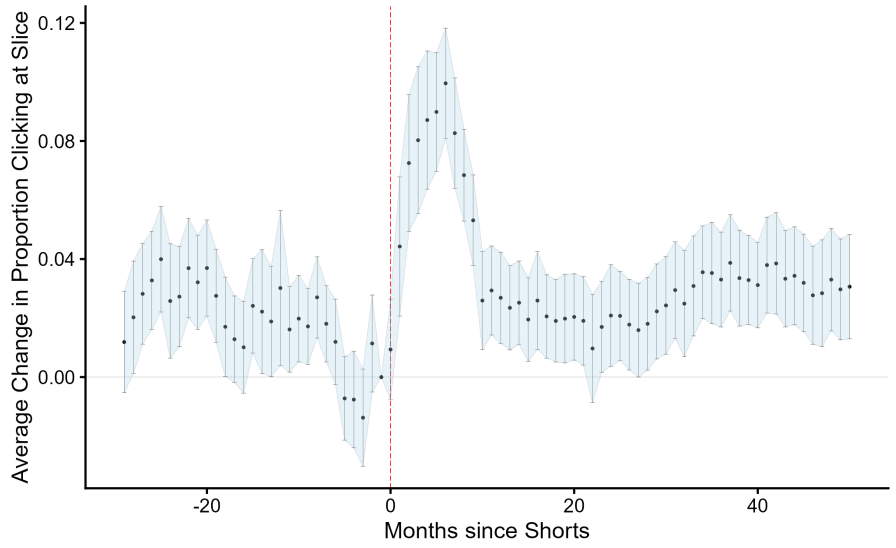
Notes: Figure 18 estimates Equation 9 separately for left-wing (Panel A), right-wing (Panel B) and podcast (Panel C) channels for videos greater than 30 minutes. These are broadly defined, and aim to identify any clear differentials. The effect is present across all groups, but is more pronounced for right-wing channels among longer videos, suggesting the audience most susceptible to *Shorts* varies by political leaning. Confidence intervals are clustered at the video level, at 95% coverage.

Figure 19: Traditional Vs Alternative Media

Panel A: Traditional Media



Panel B: Alternative Media



Notes: Figure 19 estimates Equation 9 separately for traditional media (Panel A) and alternative media (Panel B). The effect is present in both and is similar on impact. However, alternative media hosts more long form content which is affected much more overall, so this figure should be understood within that wider context. Confidence intervals are clustered at the video level, at 95% coverage.

Appendix

A Institutional Background **63**

B Additional Results and Robustness **66**

B.1 Data and Descriptives 66

B.2 Robustness: Across Channels Before and After 71

B.3 Robustness: US and India 78

B.4 Robustness: Duration 83

B.5 Robustness: Within Videos 87

B.6 Supply Side and Substitution 96

B.7 Alternative Explanations 101

C Theoretical Extensions **105**

C.1 Second Order Conditions 105

C.2 Dynamics 106

C.3 Learning about Quality 107

C.4 Ordering and Switching Costs 112

D Structural Model **114**

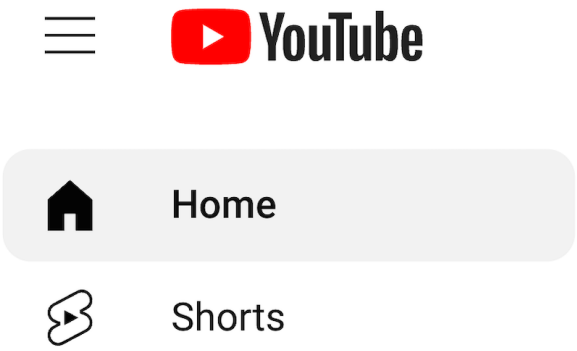
D.1 Mapping Skips to Context Value 114

D.2 Modelling Destinations 116

D.3 An Alternative Discrete-Choice Specification 117

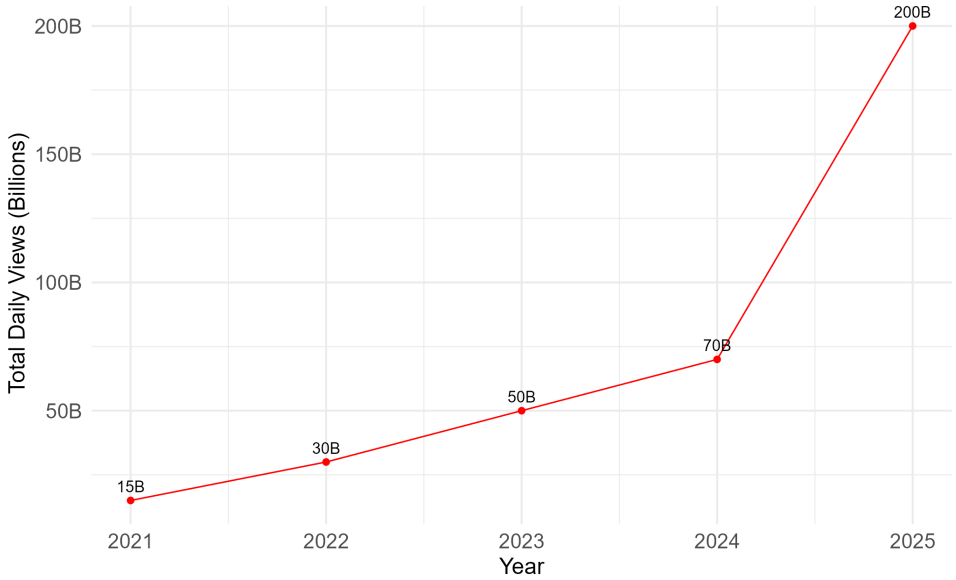
A Institutional Background

Figure A.1: Example of the *Shorts* Tab



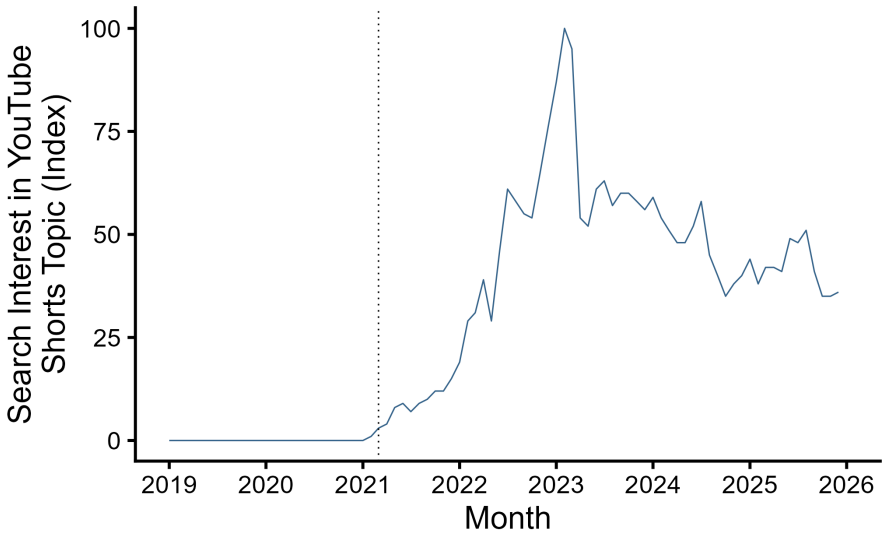
Notes: Figure A.1 shows an example of the *Shorts* tab on *YouTube*, the main method by which users access short-form content on the platform.

Figure A.2: Growth of YouTube Shorts Over Time



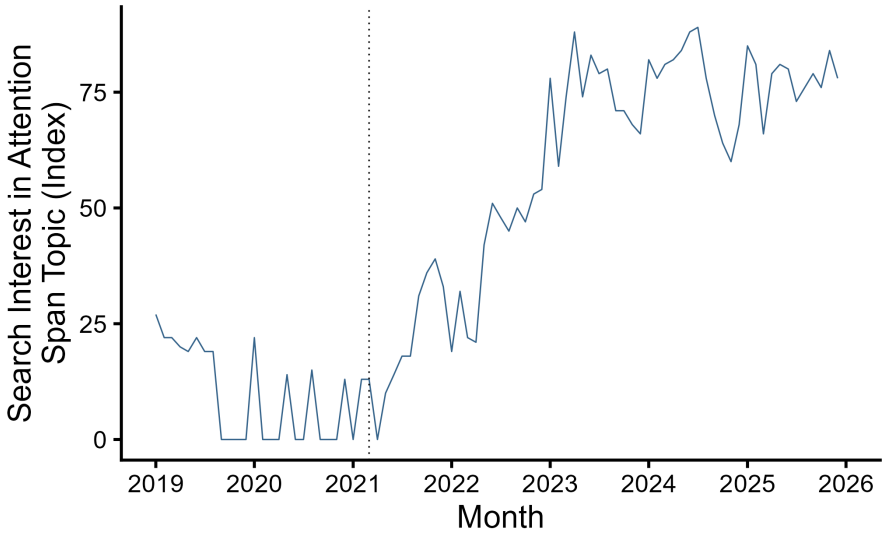
Notes: Figure A.2 shows the growth in popularity of *YouTube Shorts* platform-wide over time. *Shorts* grows substantially in popularity, while the measured effect on skipping does not, suggesting the results are not driven by re-shuffling of user composition towards low-attention-span viewers. Data taken from Teleprompter, *YouTube Shorts* statistics.

Figure A.3: Search Interest in the *YouTube Shorts* Topic



Notes: Figure A.3 plots monthly Google Trends search interest in the *YouTube Shorts* topic on YouTube search in the United States, from 2019 to 2025. The series is an index scaled so that its highest month in the window equals 100. The dotted vertical line marks March 2021, when *Shorts* launched in the US. Search interest is close to zero before the launch and rises sharply afterwards, which serves as a sanity check on the timing of the policy.

Figure A.4: Search Interest in the Attention Span Topic

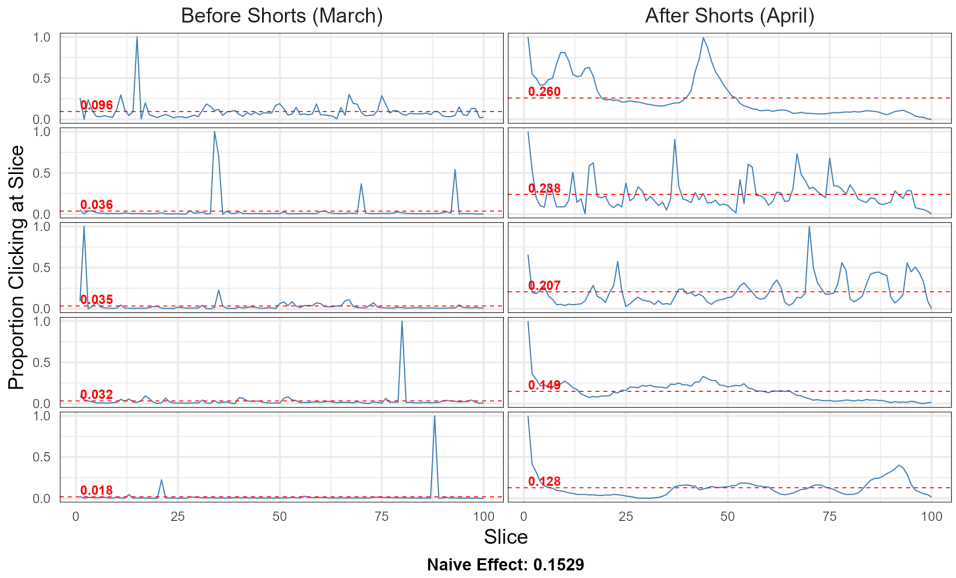


Notes: Figure A.4 plots monthly Google Trends search interest in the attention span topic on YouTube search in the United States, from 2019 to 2025. The series is an index scaled so that its highest month in the window equals 100, so levels are not comparable with Figure A.3. The dotted vertical line marks March 2021, when *Shorts* launched in the US. Search interest rises after the launch and stays high through 2025, suggesting that attention span became more salient for YouTube users. This is correlational evidence only, and may simply reflect users worrying that *Shorts* could have this effect. The sustained increase, however, is consistent with such an effect being a possibility.

B Additional Results and Robustness

B.1 Data and Descriptives

Figure B.1: Example Heatmaps Before and After the Policy: Videos Greater than 30 Minutes



Notes: Figure B.1 repeats the exercise in Figure 2 restricted to videos greater than 30 minutes. As before, heatmaps are taken from *Ben Shapiro*'s videos specifically to ensure a reasonable comparison, from the first week of March and the first week of April. The red line displays the average proportion skipping at a slice, and there is a clear increase, driven by the stark change in how videos seem to be consumed.

Figure B.2: Example Video Transcript

"more Brits than ever now say brexit was the wrong thing to do according to a new yougov survey if the brexit vote were to be held again today 55 percent would vote to remain part of the EU now this is not 75 80 85 percent it's not overwhelming but it is a definitive change uh from where it was at some point in time what do you make of this what has changed in the world that has made so many Brits reconsider let me know in the comments..."

— **The David Pakman Show**

Notes: Figure B.2 shows an excerpt of an example *YouTube*-generated transcript. In general, they are sensible and correct.

Table B.1: LDA Topics: Top 20 Terms

Panel A: Topics 1–5

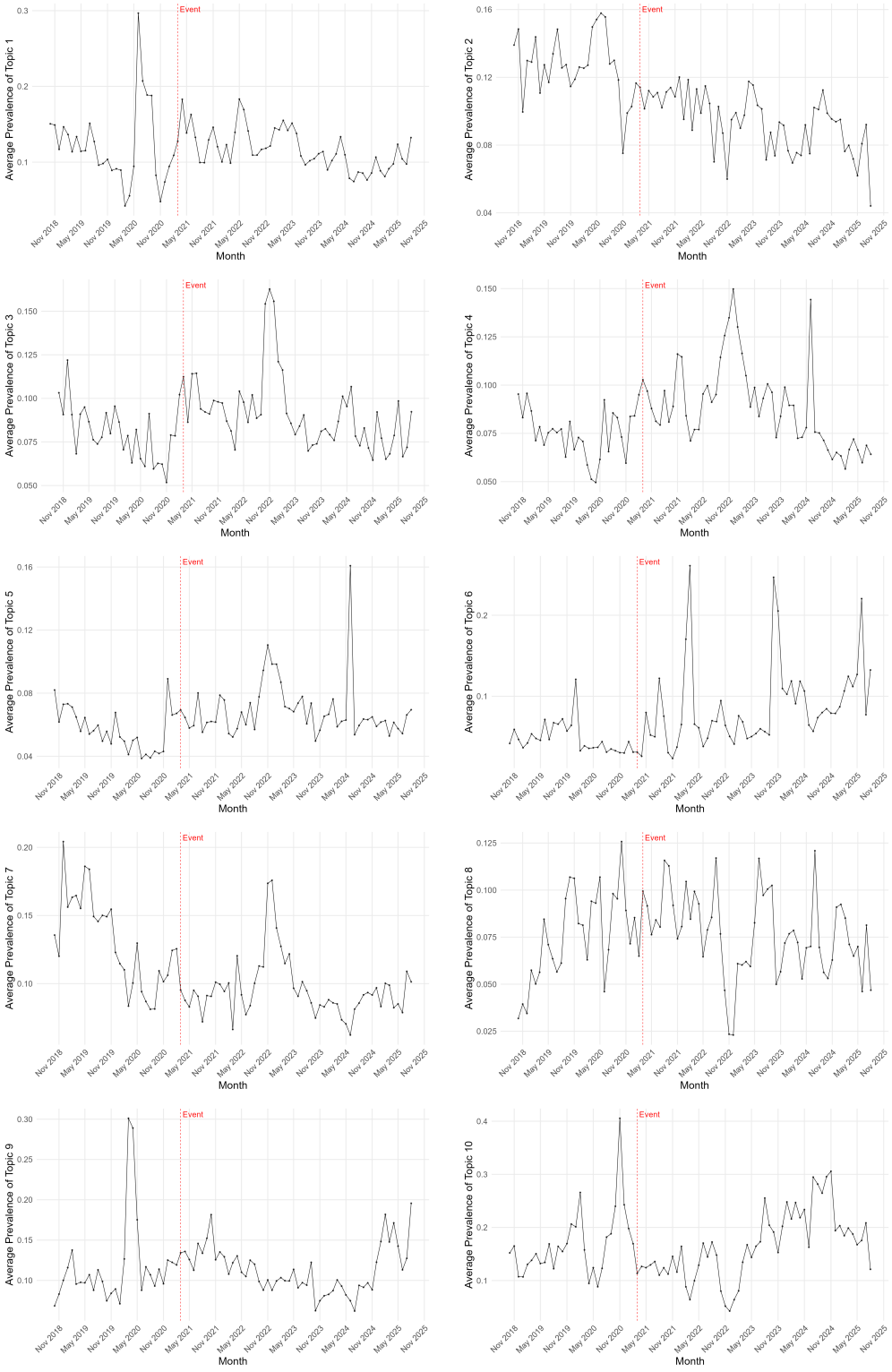
	Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
Rank	Term	$\hat{\beta}$	Term	$\hat{\beta}$	Term	$\hat{\beta}$	Term	$\hat{\beta}$	Term	$\hat{\beta}$
1	black	0.0254	york	0.0104	human	0.0118	dude	0.0236	crazy	0.0161
2	women	0.0226	daily	0.0071	sense	0.0085	fight	0.0176	god	0.0107
3	police	0.0154	free	0.0070	book	0.0076	cuz	0.0075	weird	0.0105
4	woman	0.0113	cnn	0.0062	god	0.0072	night	0.0071	cuz	0.0095
5	racist	0.0092	week	0.0053	system	0.0065	bro	0.0069	wow	0.0083
6	kids	0.0088	politics	0.0052	experience	0.0052	damn	0.0067	body	0.0076
7	children	0.0073	michael	0.0049	level	0.0050	hit	0.0067	eat	0.0074
8	school	0.0072	book	0.0048	thinking	0.0049	nice	0.0066	fun	0.0064
9	law	0.0071	ben	0.0046	ideas	0.0046	funny	0.0064	kids	0.0063
10	sex	0.0066	family	0.0044	society	0.0045	car	0.0062	amazing	0.0061
11	rights	0.0066	couple	0.0044	power	0.0043	hear	0.0056	water	0.0059
12	violence	0.0066	america	0.0043	answer	0.0042	black	0.0054	food	0.0058
13	trans	0.0065	yesterday	0.0043	science	0.0041	girl	0.0053	cool	0.0057
14	crime	0.0060	night	0.0043	read	0.0041	head	0.0052	started	0.0054
15	agree	0.0059	wire	0.0041	terms	0.0040	crazy	0.0049	super	0.0054
16	gender	0.0058	press	0.0041	mind	0.0039	listen	0.0047	podcast	0.0051
17	race	0.0058	home	0.0040	history	0.0038	fighting	0.0047	bunch	0.0051
18	racism	0.0056	truth	0.0040	conversation	0.0038	shot	0.0046	eating	0.0043
19	america	0.0053	andrew	0.0039	learn	0.0034	started	0.0043	play	0.0043
20	gay	0.0050	fox	0.0039	theory	0.0034	comedy	0.0043	imagine	0.0040

Panel B: Topics 6–10

	Topic 6		Topic 7		Topic 8		Topic 9		Topic 10	
Rank	Term	$\hat{\beta}$	Term	$\hat{\beta}$	Term	$\hat{\beta}$	Term	$\hat{\beta}$	Term	$\hat{\beta}$
1	war	0.0294	twitter	0.0110	biden	0.0461	million	0.0083	donald	0.0225
2	israel	0.0256	wait	0.0108	joe	0.0335	health	0.0080	democrats	0.0186
3	russia	0.0143	youtube	0.0084	united	0.0116	pay	0.0079	election	0.0159
4	united	0.0133	content	0.0064	administration	0.0111	dollars	0.0078	vote	0.0154
5	military	0.0132	blah	0.0063	policy	0.0087	tax	0.0068	republicans	0.0143
6	ukraine	0.0127	read	0.0060	ben	0.0074	economy	0.0065	party	0.0124
7	america	0.0104	hold	0.0058	shapiro	0.0071	companies	0.0063	court	0.0116
8	iran	0.0091	conversation	0.0056	fbi	0.0066	business	0.0062	republican	0.0112
9	china	0.0085	agree	0.0055	biden's	0.0063	market	0.0058	trump's	0.0085
10	anti	0.0071	debate	0.0046	democrats	0.0060	system	0.0057	democratic	0.0080
11	jews	0.0071	game	0.0045	quote	0.0056	company	0.0057	biden	0.0078
12	countries	0.0070	stream	0.0045	border	0.0053	jobs	0.0052	win	0.0075
13	russian	0.0065	information	0.0045	obama	0.0052	buy	0.0046	law	0.0074
14	nuclear	0.0065	post	0.0044	yesterday	0.0050	social	0.0045	campaign	0.0062
15	putin	0.0064	fine	0.0044	americans	0.0049	economic	0.0045	harris	0.0059
16	west	0.0062	opinion	0.0043	hunter	0.0049	billion	0.0044	voters	0.0058
17	foreign	0.0061	dude	0.0043	apparently	0.0047	percent	0.0044	judge	0.0057
18	jewish	0.0059	chat	0.0041	investigation	0.0046	job	0.0043	voting	0.0057
19	support	0.0049	hate	0.0041	national	0.0045	change	0.0043	supreme	0.0054
20	peace	0.0048	fake	0.0040	federal	0.0045	data	0.0043	bernie	0.0053

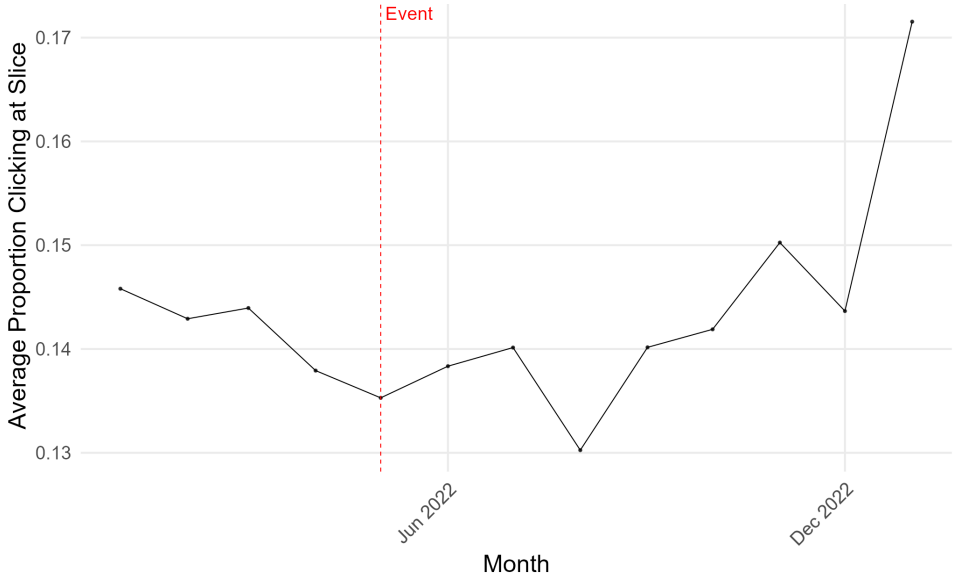
Notes: Table B.1 shows the top 20 terms by loading $\hat{\beta}$ for each of the 10 topics generated by the LDA model, which are used to flexibly control for video content.

Figure B.3: Topic Prevalence Over Time



Notes: Figure B.3 plots the average proportion of Topic x over time, with the introduction of *Shorts* shown with the red line. As we can see, there does not seem to be a contemporaneous mean shift in topics around the time of the policy.

Figure B.4: No Change in Measured Clicking Behaviour at 2022 Release

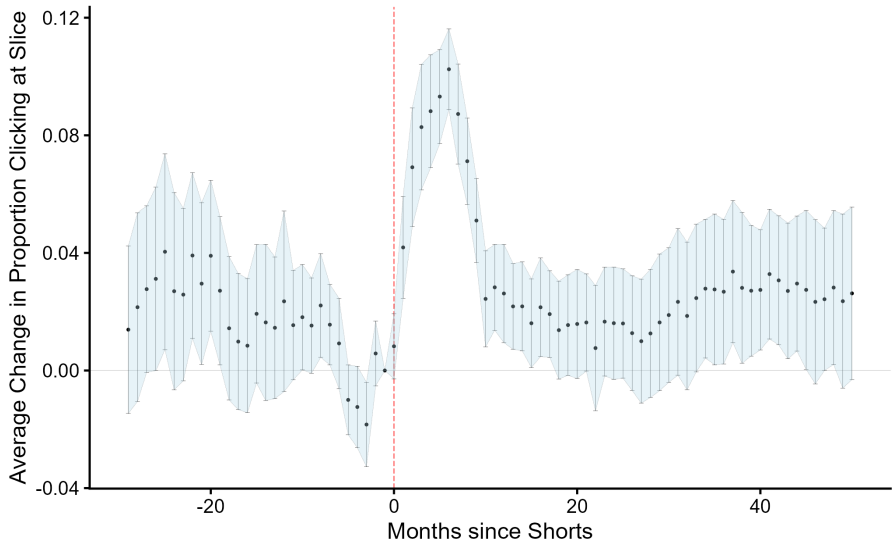


Notes: Figure B.4 shows that there is no change in measured clicking behaviour around 2022, when the heatmap was first shown to users via the video progress bar. If users responded to seeing the heatmap by clicking towards the highest spikes, we would expect a shift in measured activity at this point; the absence of such a shift supports the assumption that the heatmap reflects activity at time of upload rather than being contaminated by users reacting to the measure itself.

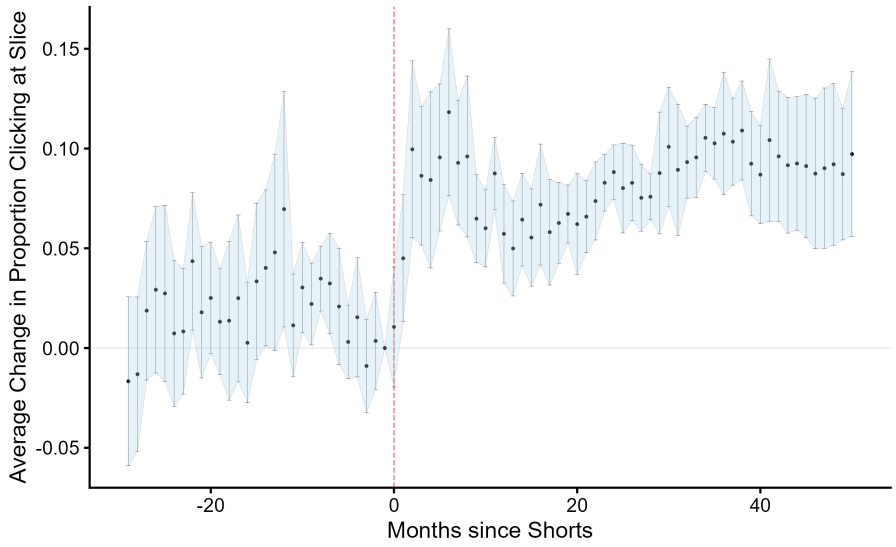
B.2 Robustness: Across Channels Before and After

Figure B.5: Attention Across Videos: Robustness to Clustering

Panel A: All Videos

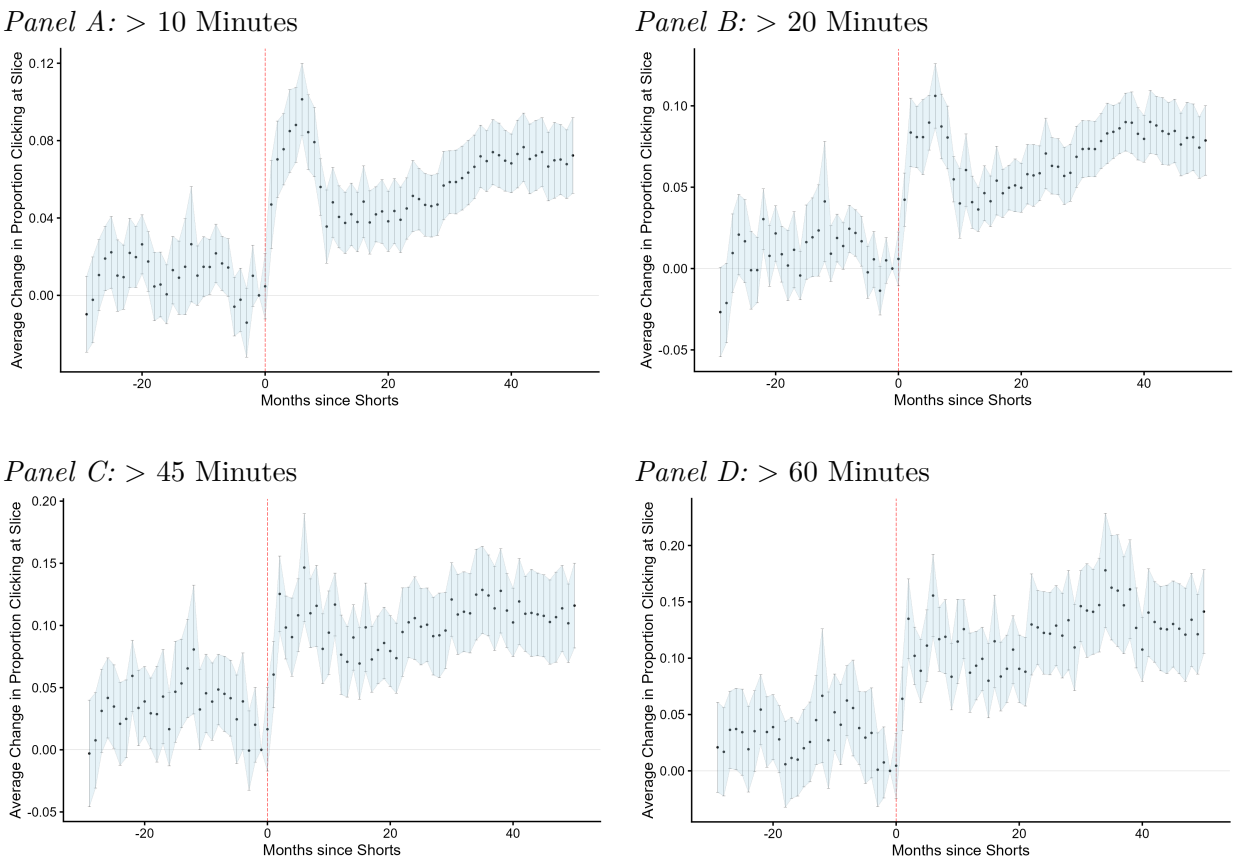


Panel B: Videos > 30 Minutes



Notes: Figure B.5 reproduces Figure 5, clustering confidence intervals at the uploader level instead of the video level. The results are materially unchanged, confirming the main result is not an artefact of within-video autocorrelation in standard errors.

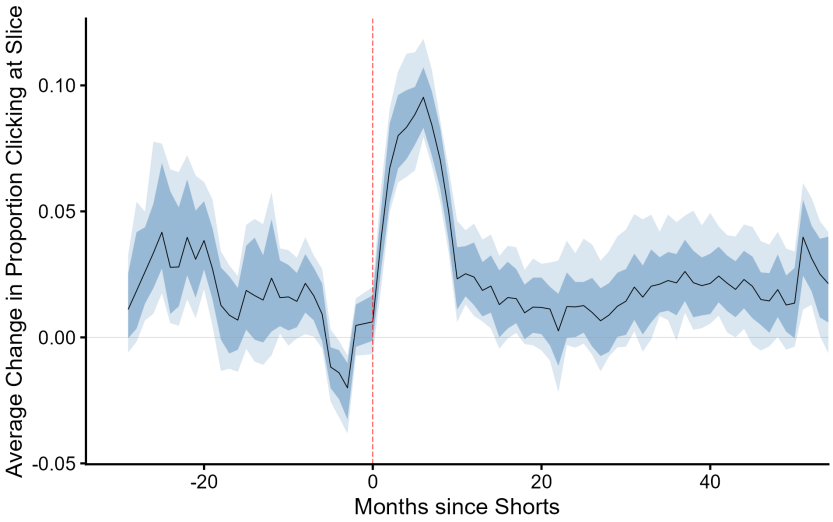
Figure B.6: Attention Across Videos: Robustness to Alternative Length Cutoffs



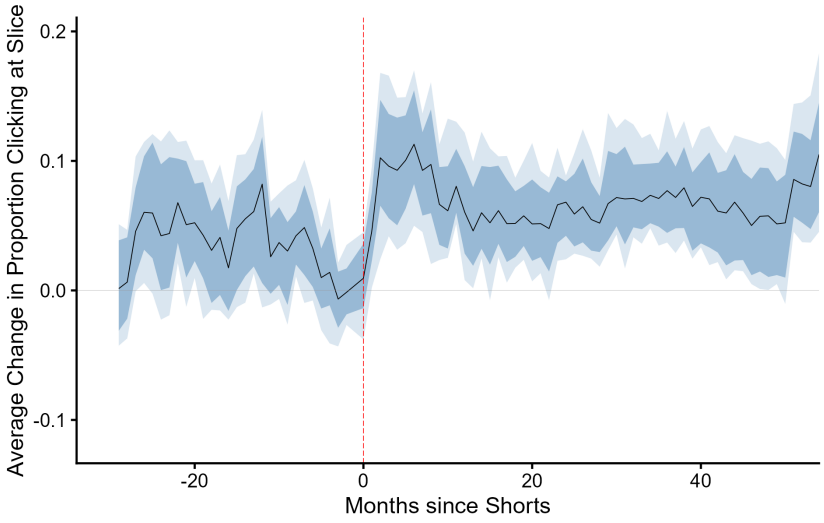
Notes: Figure B.6 reproduces Figure 5 using alternative video-length cutoffs of 10, 20, 45 and 60 minutes. The effect is present and grows monotonically with the length cut-off, confirming the choice of a 30-minute threshold elsewhere in the paper is not driving the result. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.7: Attention Across Videos: Robustness to Subsamples

Panel A: All Videos



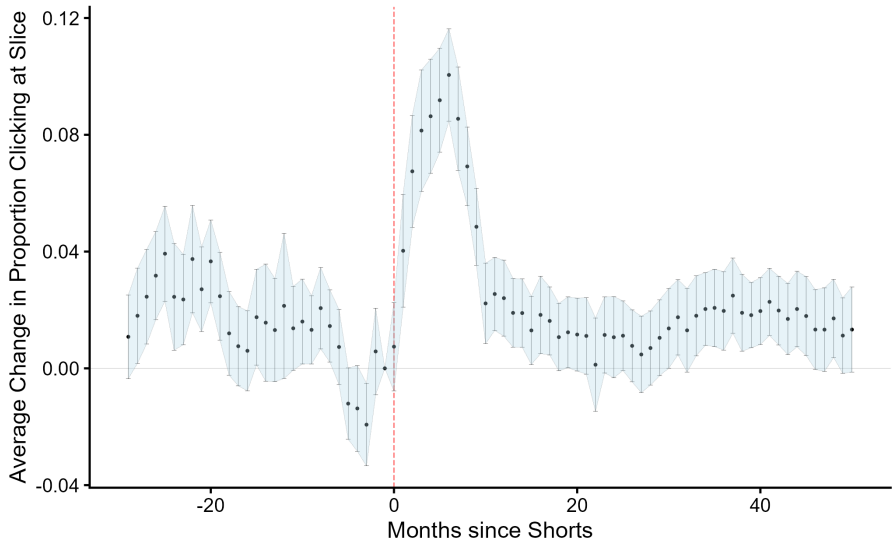
Panel B: Videos > 30 Minutes



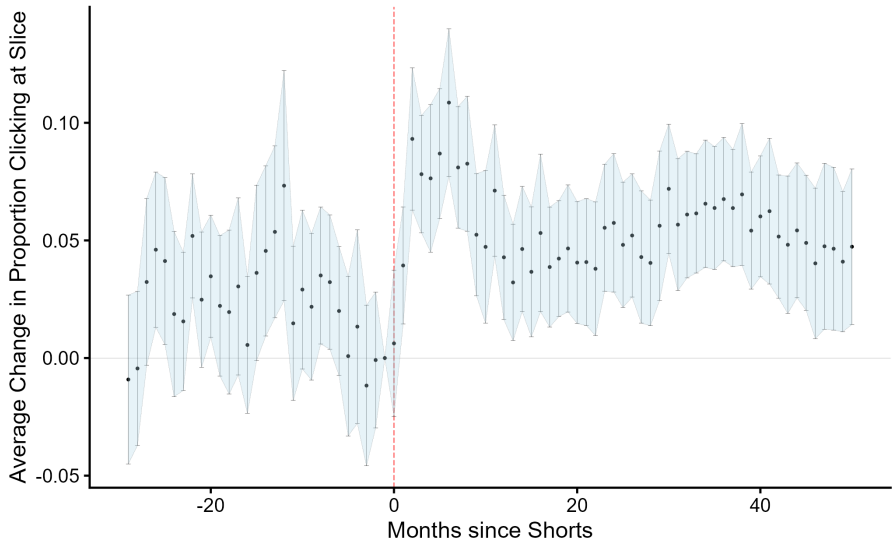
Notes: Figure B.7 reproduces Figure 5 on a 50% subsample of channels. Panel A and Panel B randomly select half of the 25 channels (rounded down) and re-estimate Equation 9. Confidence intervals are not displayed: rather the maximum deviation across these subsamples is shown in light blue, and the 95% greatest deviation in dark blue. 100 subsamples are taken in both cases, and the point estimates are calculated as usual. As we can see, there is clearly variation when some channels are dropped, but in general the results stay consistent in shape and magnitude. This strongly suggests results are not driven by specific channels.

Figure B.8: Attention Across Videos: Robustness to No Controls

Panel A: All Videos



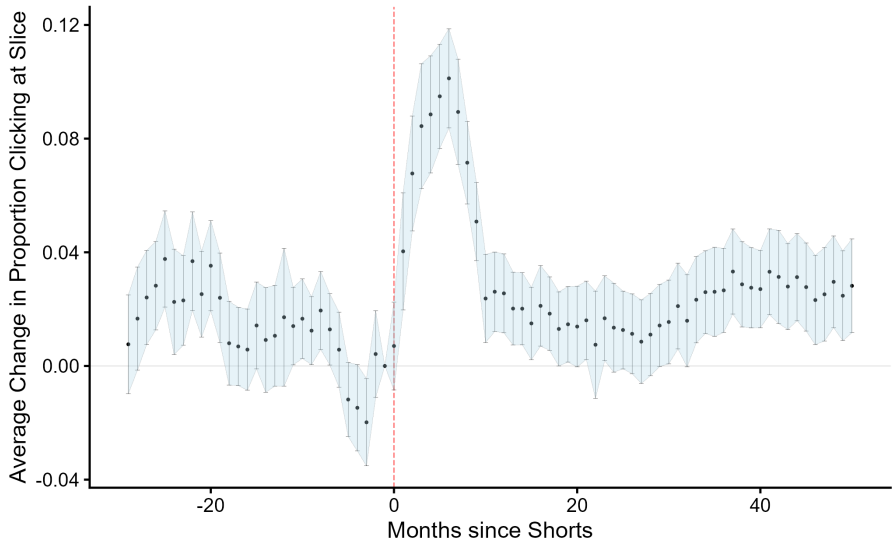
Panel B: Videos > 30 Minutes



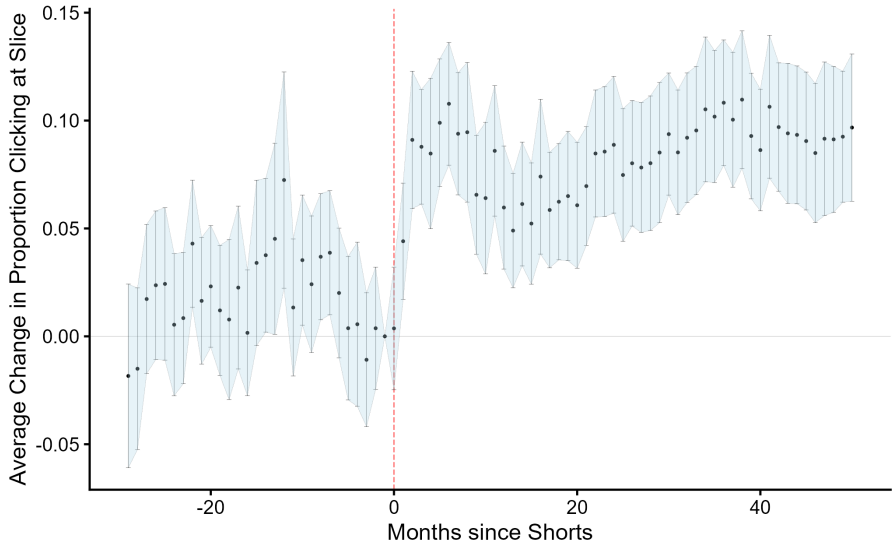
Notes: Figure B.8 reproduces Figure 5 without any additional controls for views, likes or duration. Results are close to the baseline specification, indicating the effect is not sensitive to the inclusion of flexible controls. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.9: Attention Across Videos: Robustness to Topic Fixed Effects

Panel A: All Videos

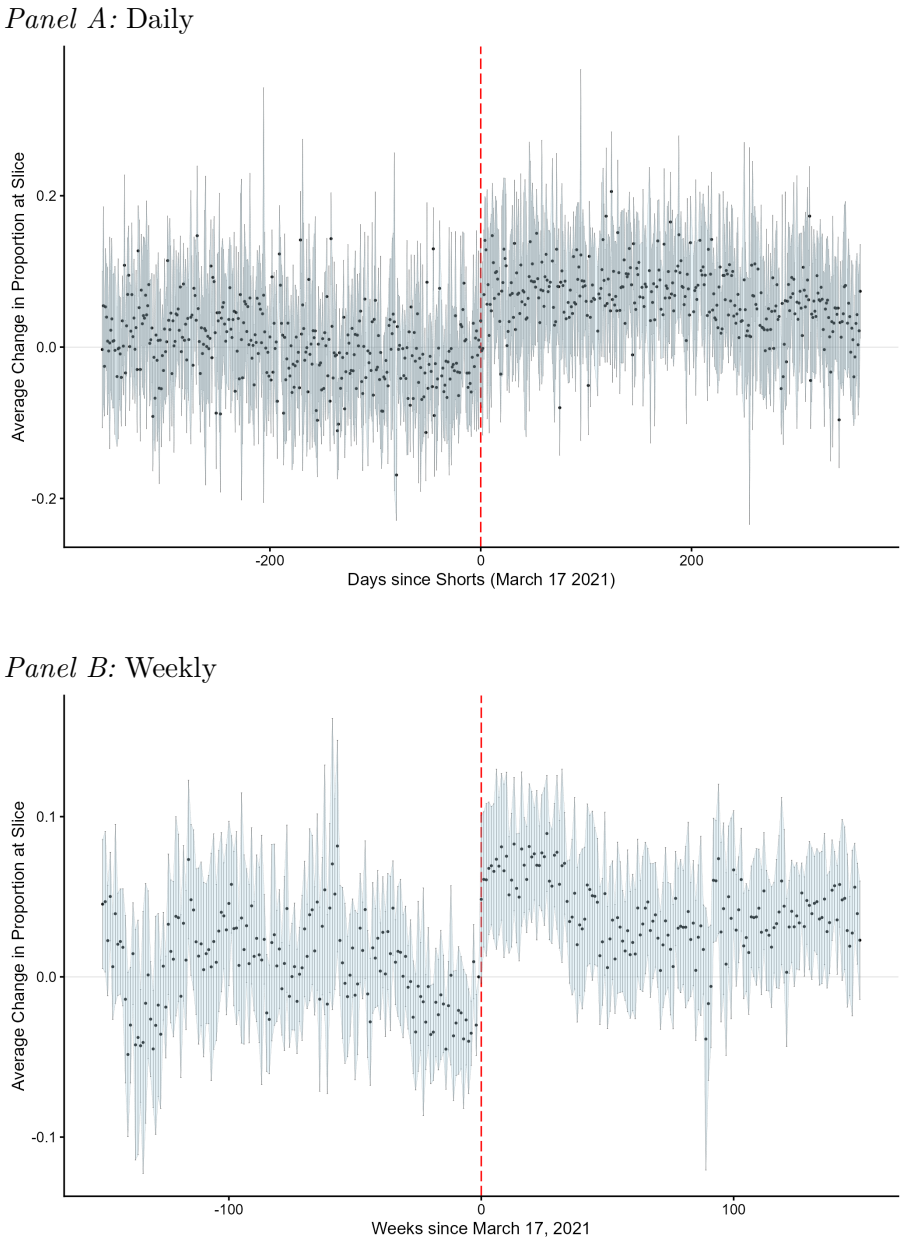


Panel B: Videos > 30 Minutes



Notes: Figure B.9 reproduces Figure 5 including dominant topic fixed effects. The effect is unchanged by flexibly controlling for topic, confirming it is not driven by compositional shifts in the mix of video topics around the policy. Confidence intervals are clustered at the video level, at 95% coverage.

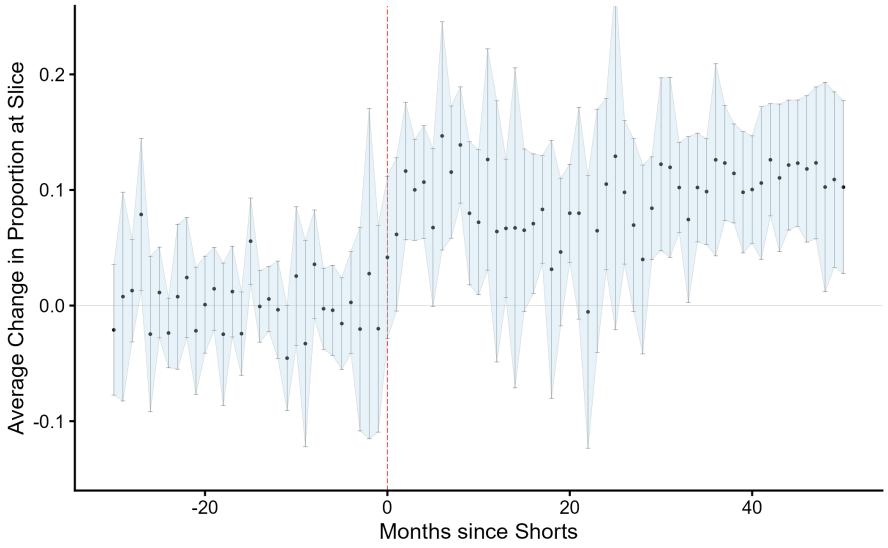
Figure B.10: Attention Across Videos: Daily and Weekly Event Studies



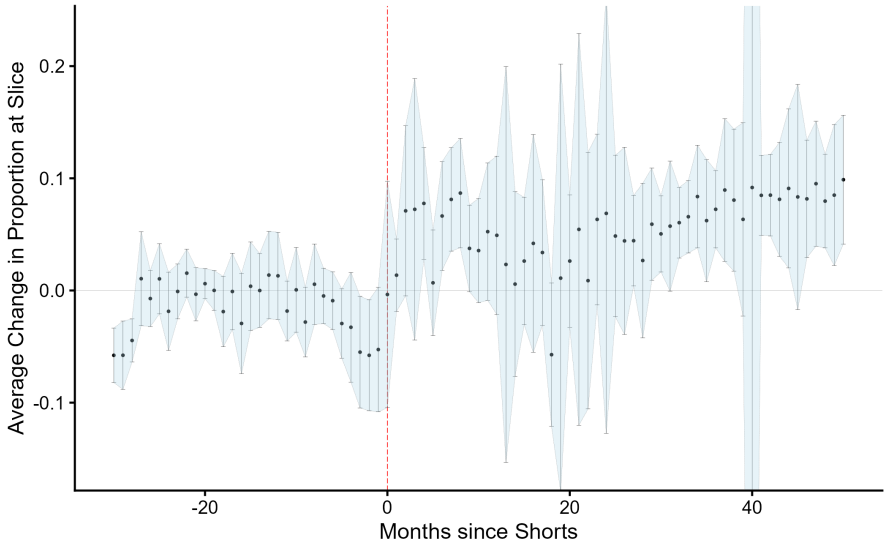
Notes: Figure B.10 re-estimates Equation 9 with a unique coefficient for each day (Panel A) and week (Panel B) since the policy, complementing the monthly-level event study in Figure 5. Both are noisier than the monthly specification, but the average increase visibly begins within the first two weeks of the policy, supporting the immediate timing of the effect. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.11: Equation 9: Alternative Estimation Strategies

Panel A: Callaway and Sant'Anna



Panel B: Borusyak, Jaravel and Spiess

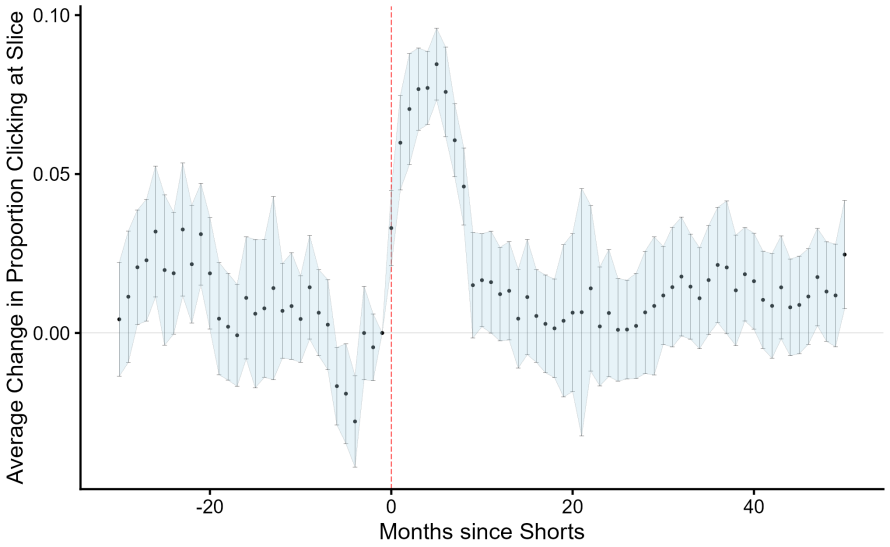


Notes: Figure B.11 recreates Figure 5 following Callaway and Sant'Anna (2021) in Panel A and Borusyak et al. (2024) in Panel B. Confidence intervals are clustered at the uploader level, at 95% coverage. In Panel B at month 40 the confidence interval is truncated to ensure readability due to a uniquely wide band .

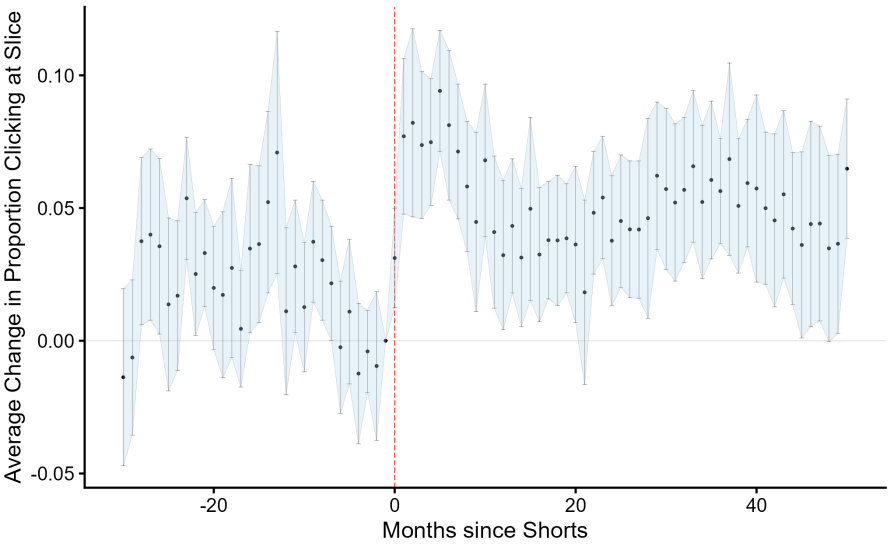
B.3 Robustness: US and India

Figure B.12: Attention Across Videos: Robustness to Clustering (Absolute)

Panel A: All Videos

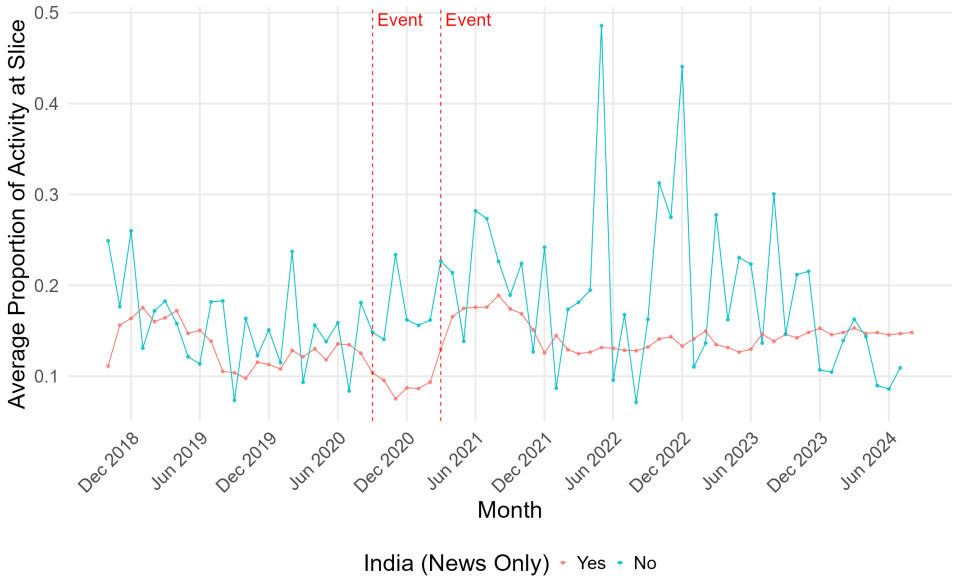


Panel B: Videos > 30 Minutes



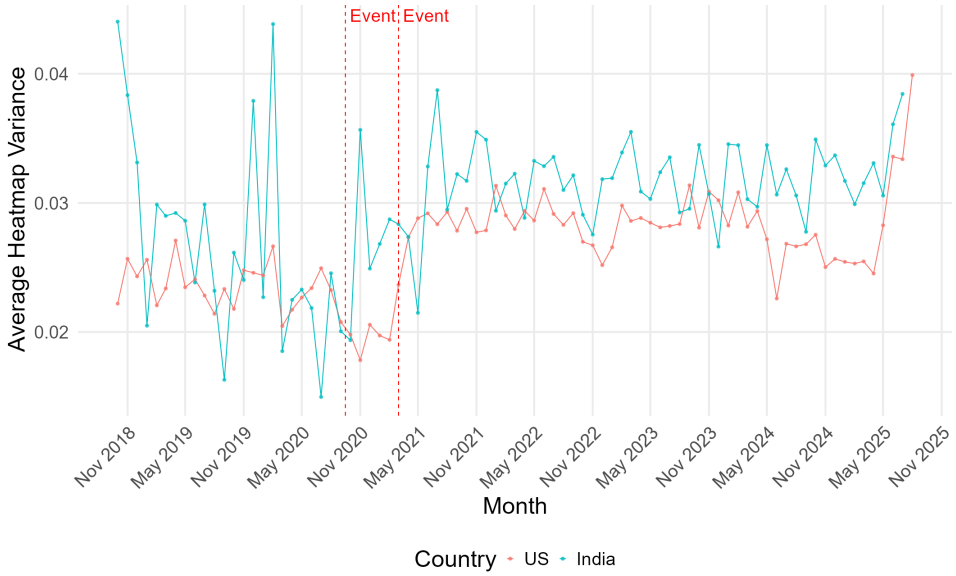
Notes: Figure B.12 reproduces Figure 6, clustering confidence intervals at the uploader level instead of the video level. Results are unchanged, confirming the India comparison is not sensitive to the level at which standard errors are clustered.

Figure B.13: DiD Comparing US and India: Raw Data, News Channels Only



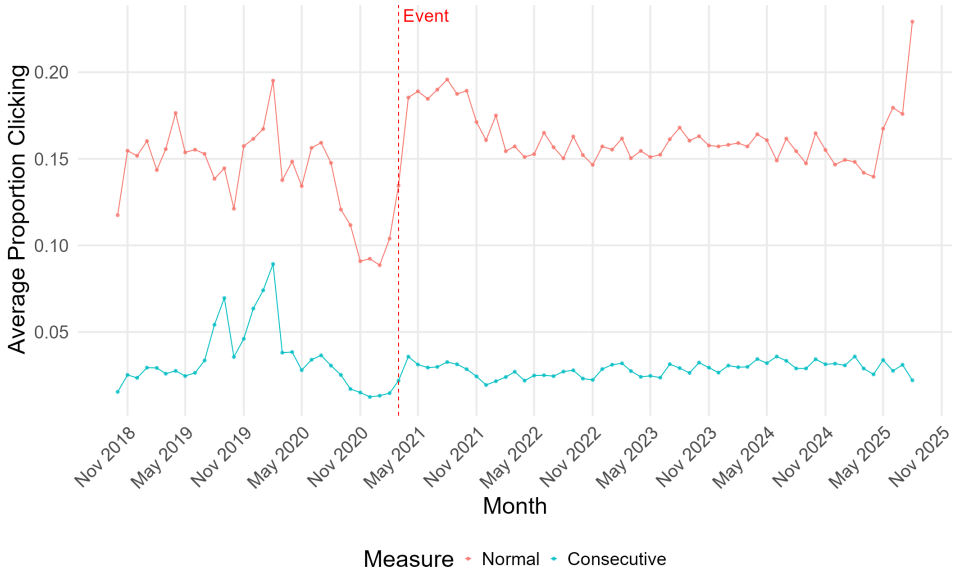
Notes: Figure B.13 repeats Figure 7 restricting the Indian sample to the three largest Indian news channels (*AajTak*, *IndiaTV* and *ABP News*), dropping the two broader media channels. Additionally this is not a rolling average, so the series are noisier, but the two mean shifts remain clearly visible and correctly timed. Skipping rises in India from September 2020 and in the US from March 2021, with no contemporaneous jump in the other market on either date. The pattern is therefore not driven by the inclusion of the broader media channels.

Figure B.14: DiD Comparing US and India: Heatmap Variance



Notes: Figure B.14 shows a rolling average of the average variance of a video heatmap comparing US channels to the 5 Indian channels in my sample, for videos at least 30 minutes long. As before there is importantly no contemporaneous spike in March 2021 for Indian channels, ruling out measurement changes or updates driving the observed result. Second, there seems to be a less sudden but present increase in variance when *Shorts* was released in India. *Shorts* exposure spikes the variance of attention only in India in September 2020, and only in the US in March 2021, confirming the timing of the effect tracks the staggered rollout rather than a common platform-wide shock. This strongly suggests *Shorts* causally increased skipping.

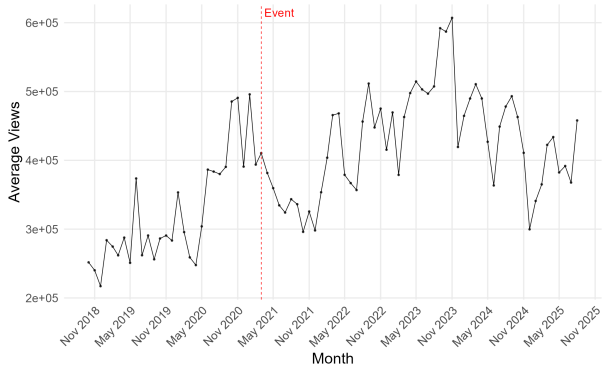
Figure B.15: Consecutive vs Non-Consecutive Activity



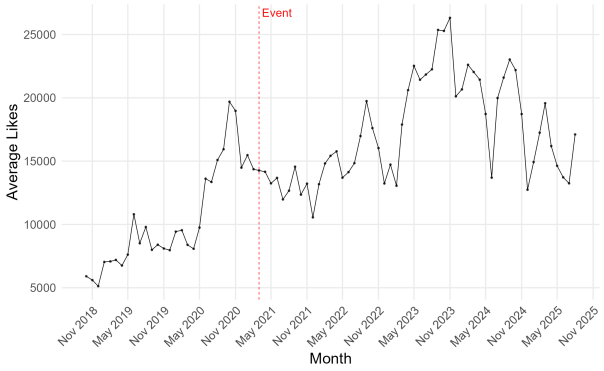
Notes: Figure B.15 shows the usual overall monthly average activity for videos greater than 30 minutes. If skipping is driving the effect, it should be driven by spikes in activity that are not spread out perfectly consecutively across slices; it seems very unlikely that skipping takes place entirely from slice N to N+1. To check this, I construct a measure of activity that can only be equal to or lower than the previous slice's activity, which abstracts from clicking through and isolates only changes that follow a smooth drop-off curve similar to Figure 4. As we can see, this completely nullifies the jump in activity, strongly supporting the click-through hypothesis as one measure changes sharply while the other does not. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.16: Views, Likes and Comments Around the Policy

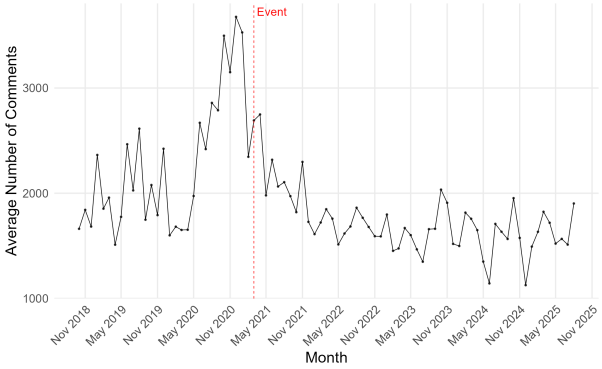
Panel A: Views



Panel B: Likes



Panel C: Comment Count

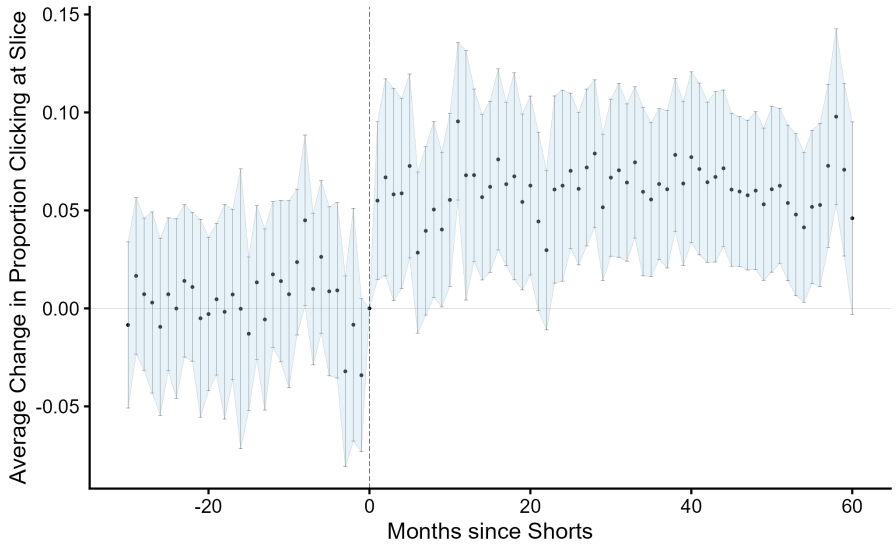


Notes: Figure B.16 shows the raw data of views (Panel A), likes (Panel B) and comment count (Panel C) around the introduction of *Shorts*. None of these measures of engagement shows a sudden change around the policy, consistent with the increase in skipping being driven by a change in existing user behaviour rather than a change in viewer composition.

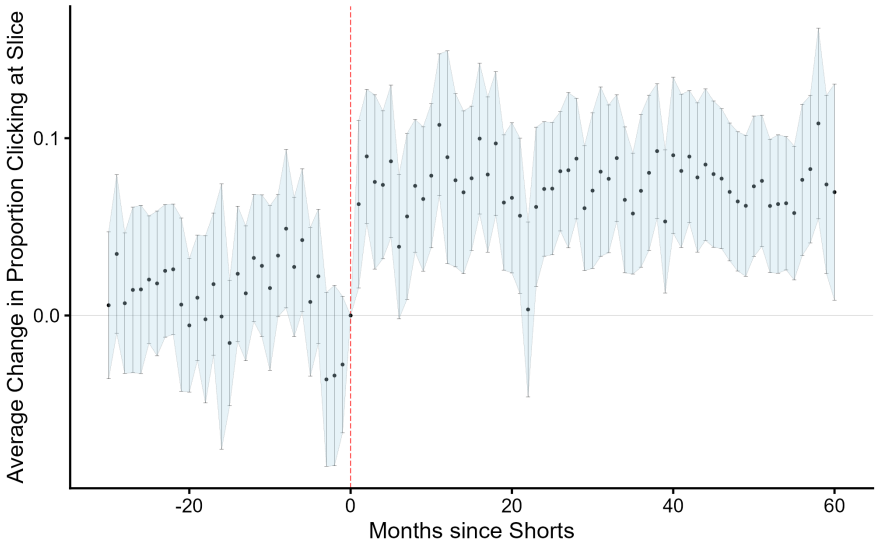
B.4 Robustness: Duration

Figure B.17: Attention Across Videos: Alternative Treatment Definitions

Panel A: Treated if $>$ 30 mins



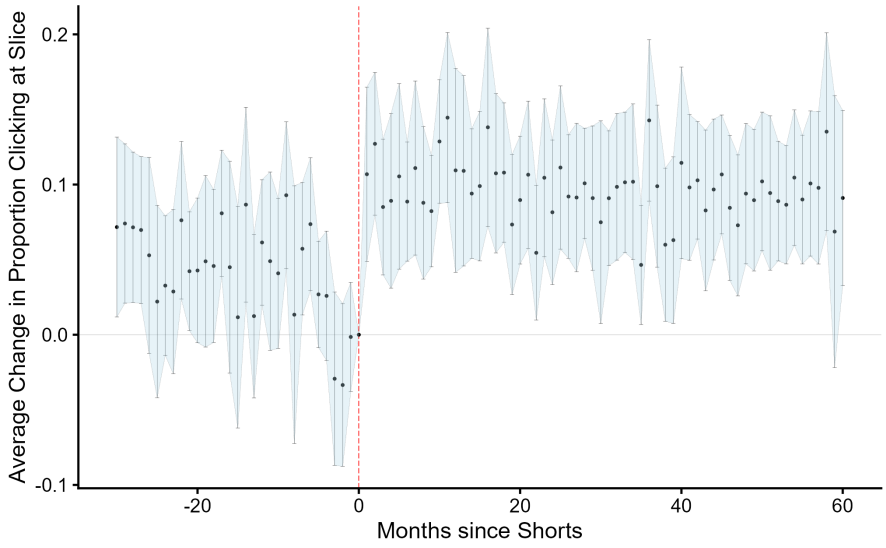
Panel B: Treated if $>$ 45 mins



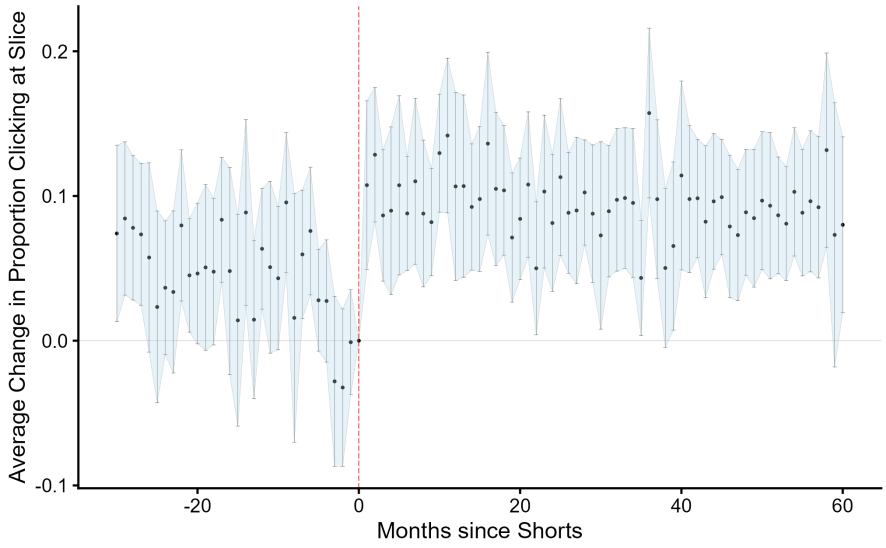
Notes: Figure B.17 compares across videos following Equation 11. Panel A defines treated and control as above and below 30 minutes respectively. Panel B defines treated and control as above and below 45 minutes respectively. Results are very similar to the 60-minute baseline in Figure 8, confirming the choice of cut-off is not driving the pattern. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.18: Attention Across Videos: Duration, Topic Robustness

Panel A: Topic Fixed Effects



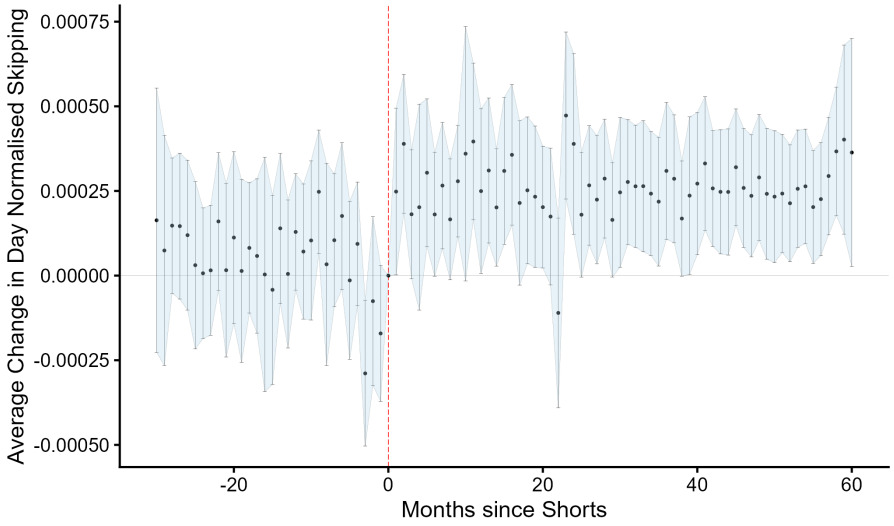
Panel B: Extensive Controls



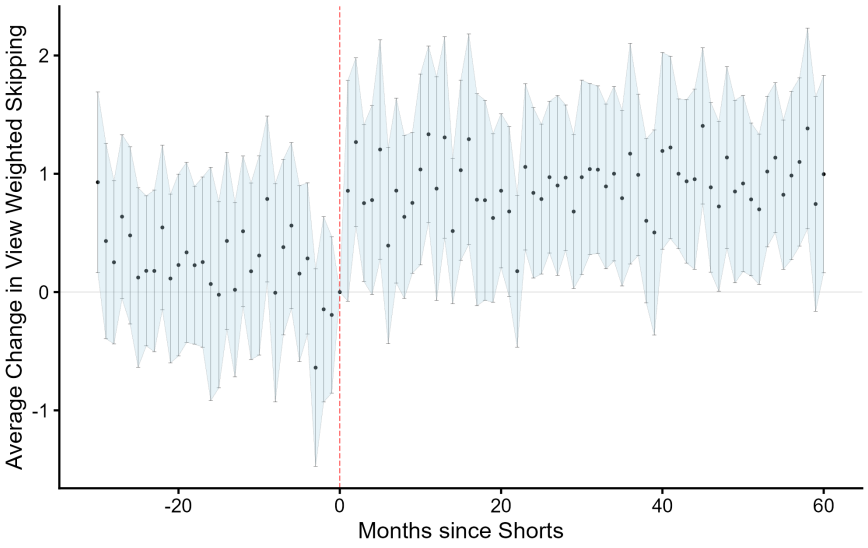
Notes: Figure B.18 reproduces Figure 8, controlling flexibly for dominant topic. Panel A includes topic fixed effects, Panel B a more extensive set of controls: views and likes × uploader × linear time. The duration effect is robust to both, confirming it is not driven by a compositional shift in the topics covered by longer videos. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.19: Attention Across Videos: Duration, Day-Normalised and View-Weighted

Panel A: Day-Normalised

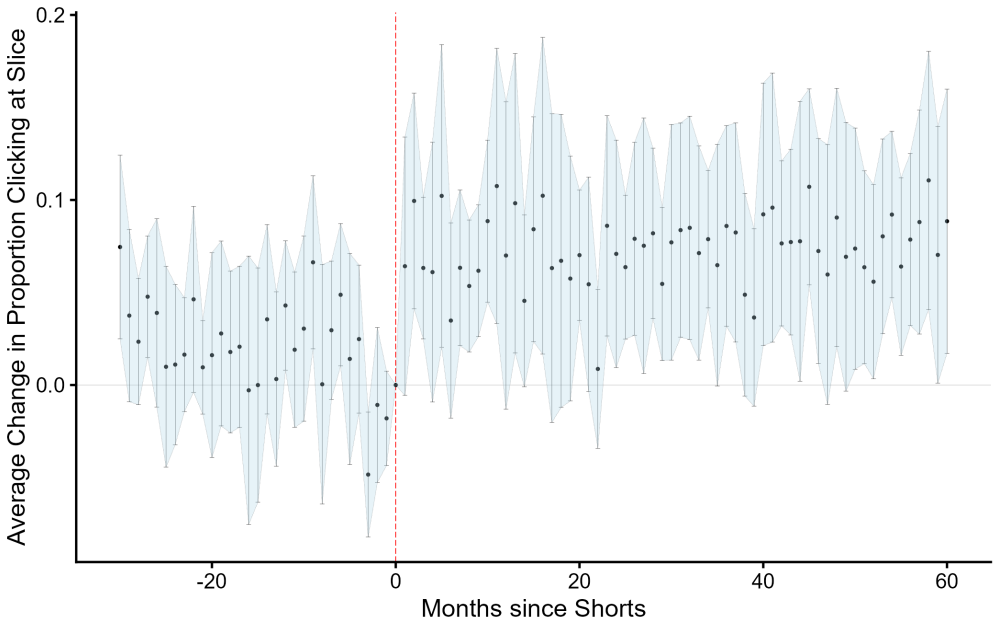


Panel B: View-Weighted



Notes: Figure B.19 reproduces Figure 8. Panel A normalises activity by day, Panel B weights observations by views. Results are similar to the baseline, indicating the effect is not an artefact of how activity is normalised or driven disproportionately by the most-viewed videos. Confidence intervals are clustered at the video level, at 95% coverage.

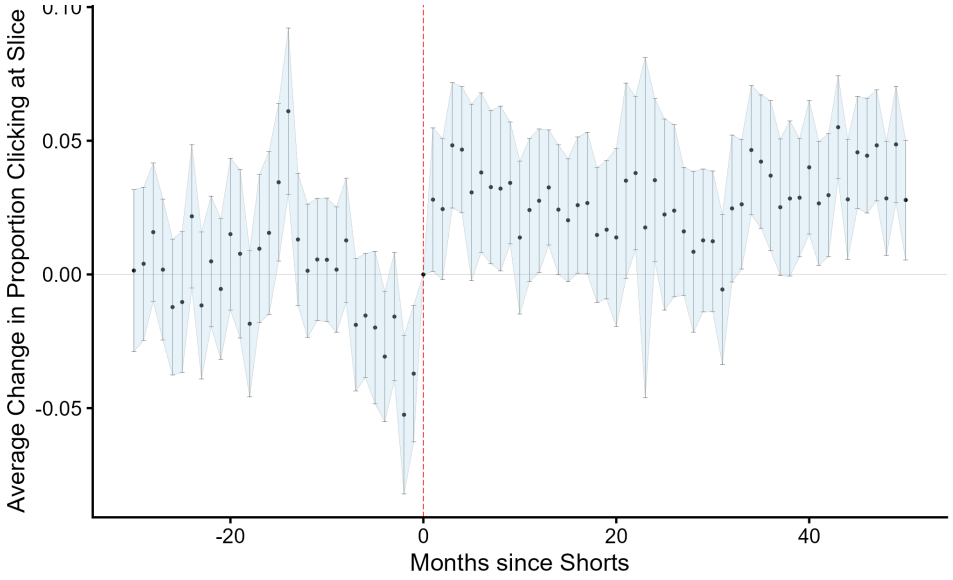
Figure B.20: Attention Across Videos: Duration, Wild Cluster Bootstrap



Notes: Figure B.20 re-estimates Equation 11 using wild cluster bootstrap with 1000 replications. This is done on the uploader level only. Results are robust to clustering at the video level and the uploader × month level, and at the uploader level. However, with 25 uploaders there could be concern that the number of clusters is too few to accurately gauge robustness. Nonetheless, the results are still significant, even at the event study level, when using wild cluster bootstrap.

B.5 Robustness: Within Videos

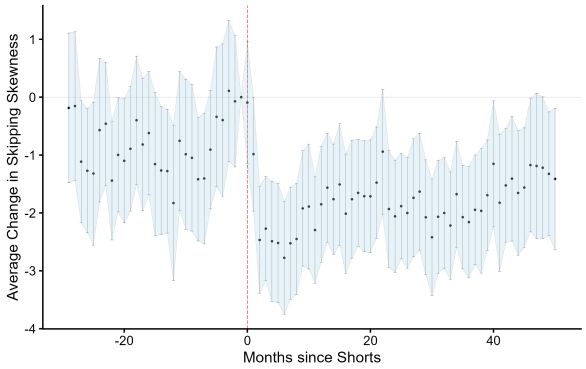
Figure B.21: Within Videos: Event Study



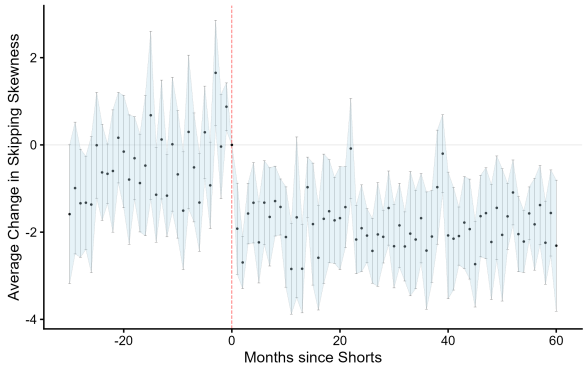
Notes: Figure B.21 estimates a difference-in-differences for videos at least 30 minutes long, comparing the body of the video to its end before and after the introduction of *Shorts*, with video fixed effects included. The body is defined as slices 2 to 95, and the end 96 to 100. There is a clear sustained increase in skipping in the body relative to the end, bolstering the causal interpretation of the within-video results and justifying the estimation strategy used for the structural model. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.22: Within-Video Spread: Skewness and Kurtosis

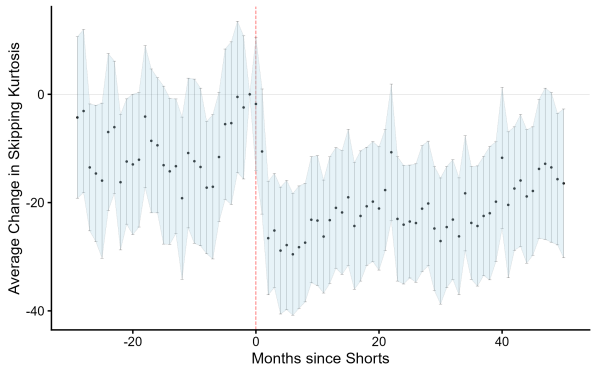
Panel A: Skewness, Overall



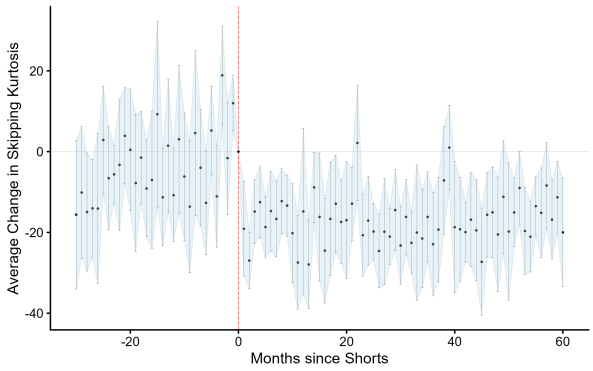
Panel B: Skewness, Long Vs Short



Panel C: Kurtosis, Overall



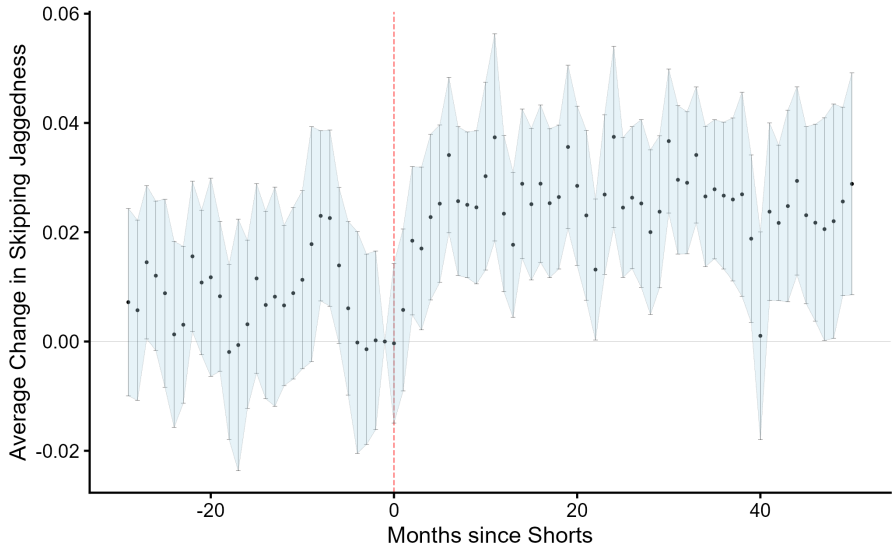
Panel D: Kurtosis, Long Vs Short



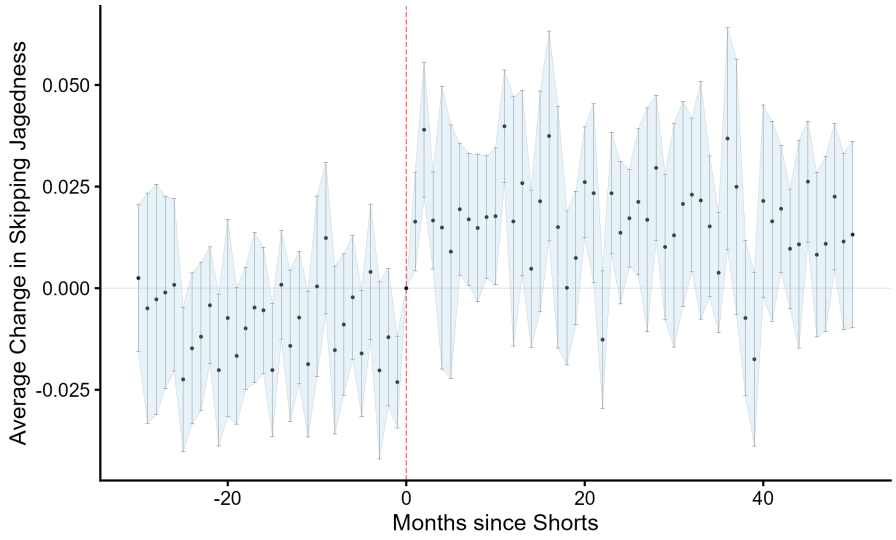
Notes: Figure B.22 shows the average skewness (Panels A-B) and kurtosis (Panels C-D) of the video-level heatmap. Panels A and C show the overall effect following Equation 9 for videos longer than 30 minutes, Panels B and D compare long and short videos following Equation 11. Both fall after the policy, consistent with the increase in skipping occurring broadly across the video rather than being concentrated at a few cut points. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.23: Within-Video Spread: Jaggedness

Panel A: Overall

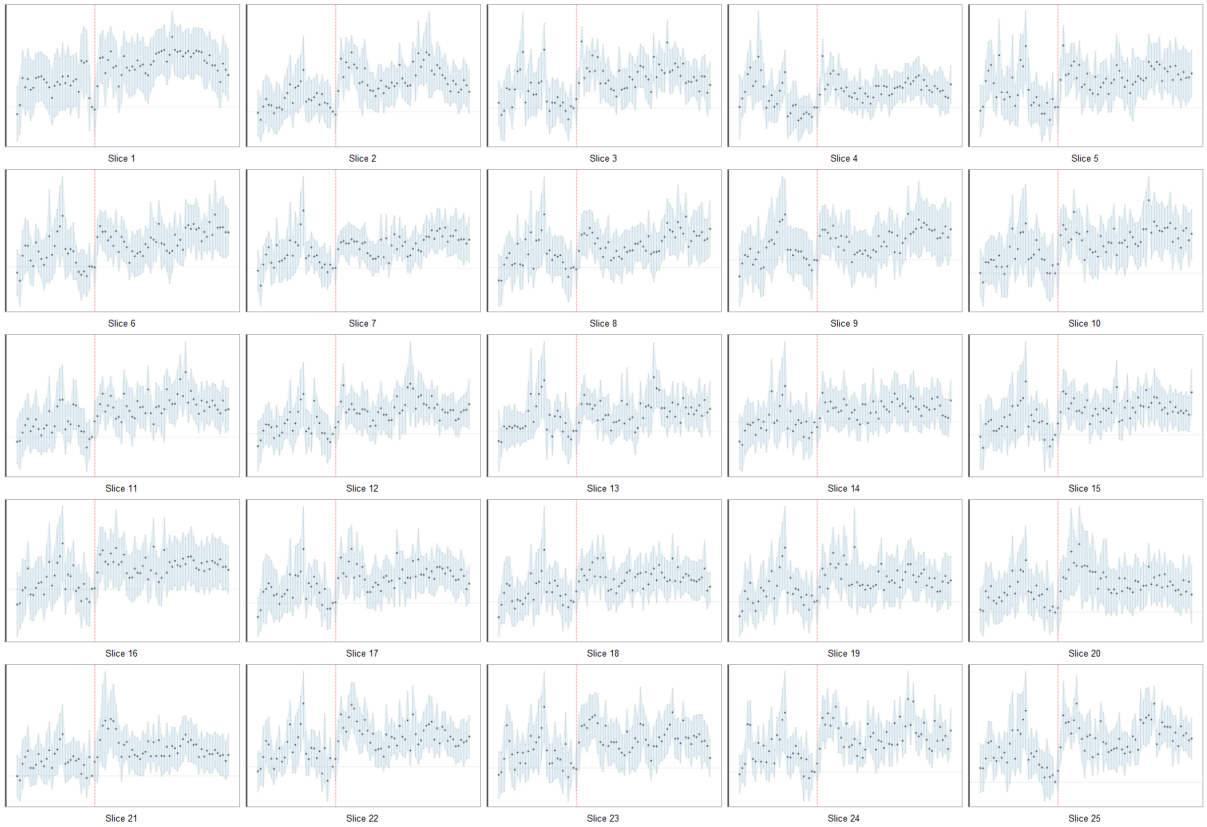


Panel B: Long Vs Short



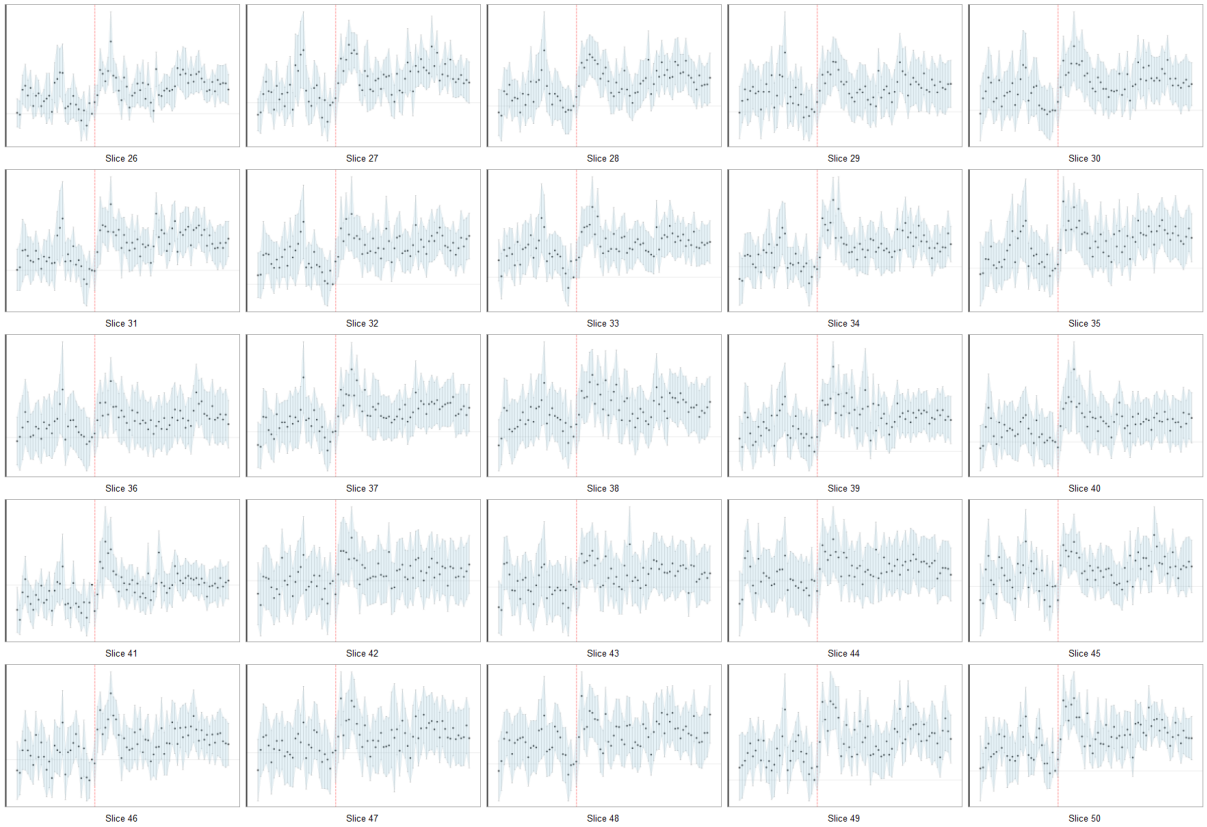
Notes: Figure B.23 shows the average absolute deviation between consecutive slices (jaggedness) of the video-level heatmap. Panel A shows the overall effect, Panel B compares long and short videos. Jaggedness increases after the policy, consistent with a greater rate of change in skipping as slices increase, even as the overall distribution of skipping becomes more uniform. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.24: Slice-Level Event Studies: Slices 1–25



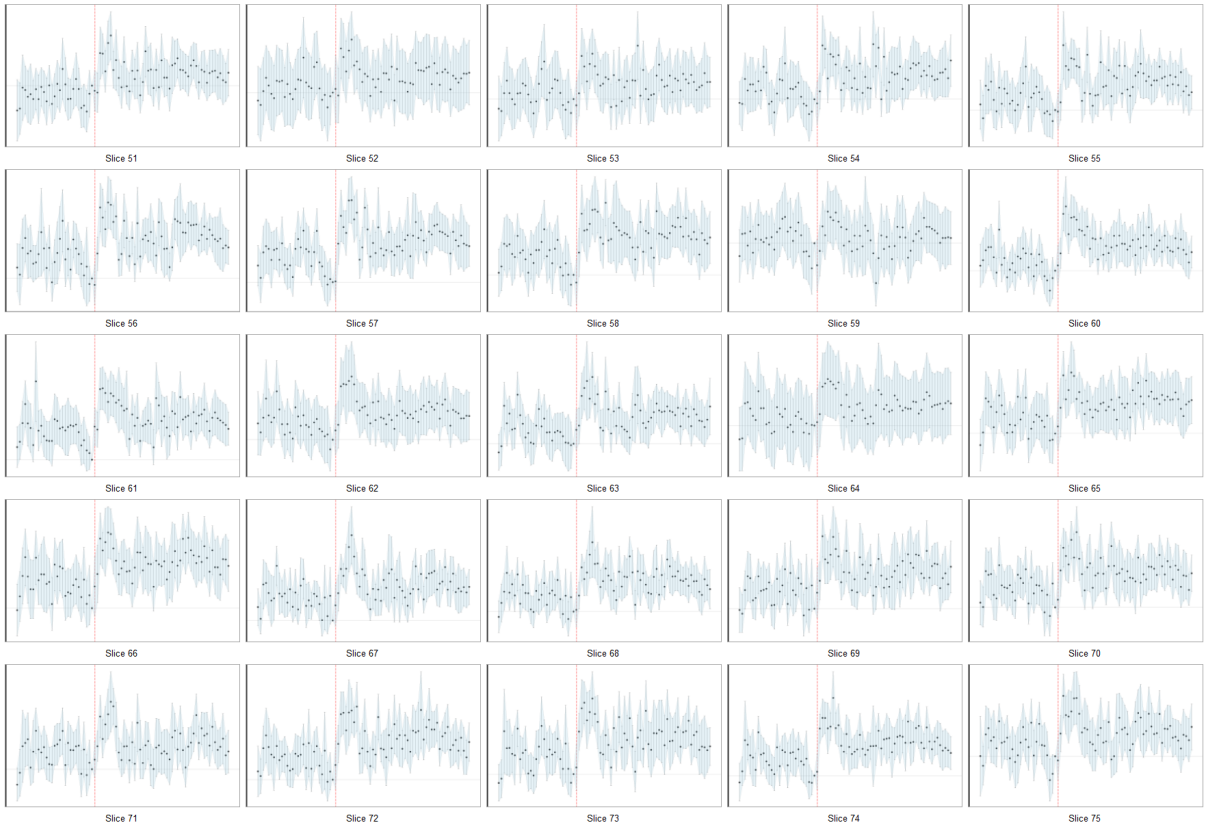
Notes: Figure B.24 shows the individual event study for each slice from 1 to 25, estimated as in Figure 11. Slice 1 shows no effect, while the remaining slices show a broadly correct mean shift, though it is noisy. The crucial take-away is that Figure 11 is a reasonable way to aggregate effects. In each panel the y axis is the average proportion clicking at a slice, and the x axis is months since *Shorts* with the red dashed line their introduction. Labels have been suppressed for readability. Confidence intervals are clustered at the uploader level, at 95% coverage.

Figure B.25: Slice-Level Event Studies: Slices 26–50



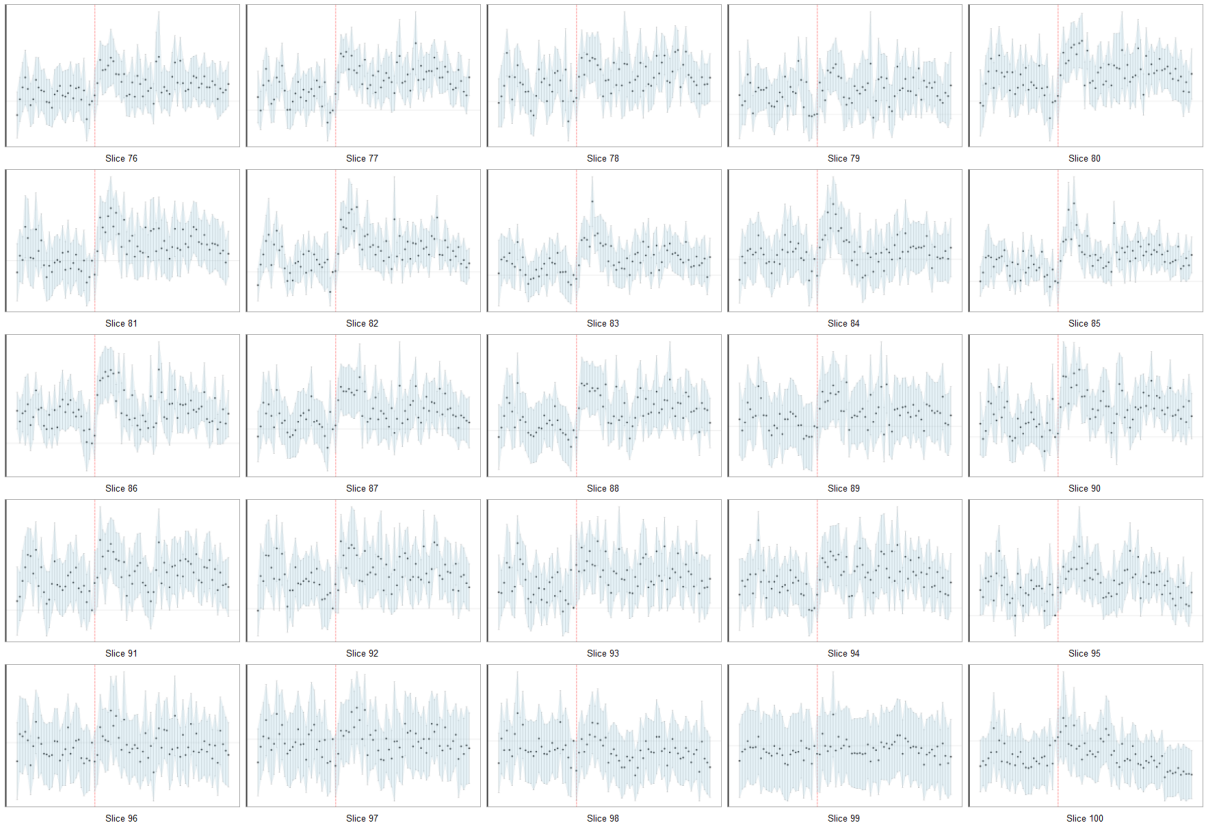
Notes: Figure B.25 shows the individual event study for each slice from 26 to 50, estimated as in Figure 11. In each panel the y axis is the average proportion clicking at a slice, and the x axis is months since *Shorts* with the red dashed line their introduction. Labels have been suppressed for readability. Confidence intervals are clustered at the uploader level, at 95% coverage.

Figure B.26: Slice-Level Event Studies: Slices 51–75



Notes: Figure B.26 shows the individual event study for each slice from 51 to 75, estimated as in Figure 11. In each panel the y axis is the average proportion clicking at a slice, and the x axis is months since *Shorts* with the red dashed line their introduction. Labels have been suppressed for readability. Confidence intervals are clustered at the uploader level, at 95% coverage.

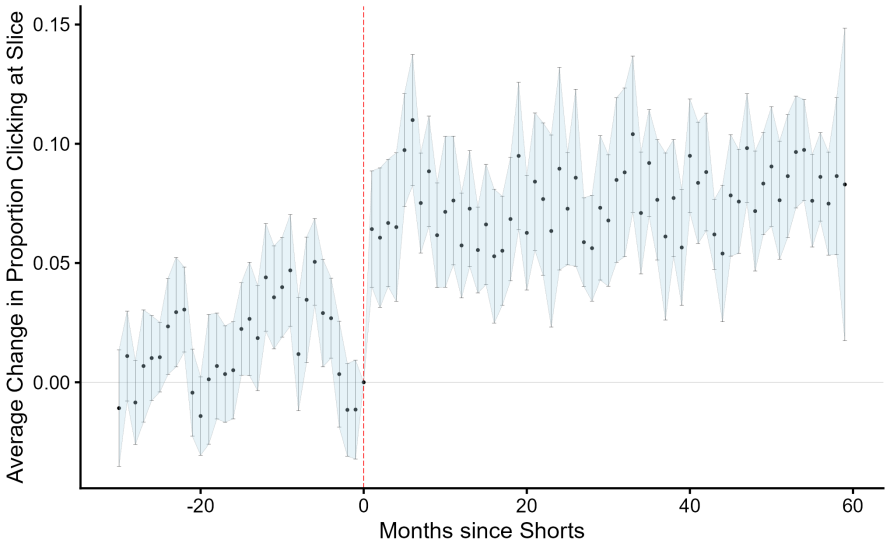
Figure B.27: Slice-Level Event Studies: Slices 76–100



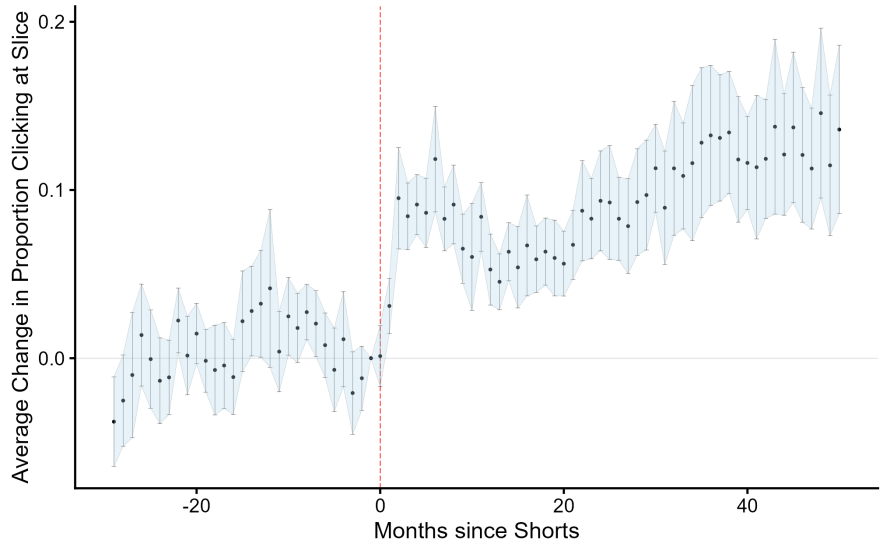
Notes: Figure B.27 shows the individual event study for each slice from 76 to 100, estimated as in Figure 11. The effect visibly drops off towards the very final slices, consistent with the null result documented at the end of a video throughout the paper. In each panel the y axis is the average proportion clicking at a slice, and the x axis is months since *Shorts* with the red dashed line their introduction. Labels have been suppressed for readability. Confidence intervals are clustered at the uploader level, at 95% coverage.

Figure B.28: Chapter Event Studies

Panel A: Chapter DiD



Panel B: Effect between Chapters



Notes: Figure B.28 examines the effect of chapter markers on skipping. Panel A shows a difference-in-differences comparing slices with a chapter marker to those without, before and after the introduction of *Shorts*, revealing an additional increase in skipping to chapter markers after the policy. Panel B re-estimates Equation 9 restricted to slices that do not contain a chapter marker, for videos which are split into chapters. Skipping still increases in Panel B, showing that the rise in skipping occurs *within* chapters and not merely at chapter boundaries, which is inconsistent with users simply splitting videos at the points the uploader provides. Confidence intervals are clustered at the video level, at 95% coverage.

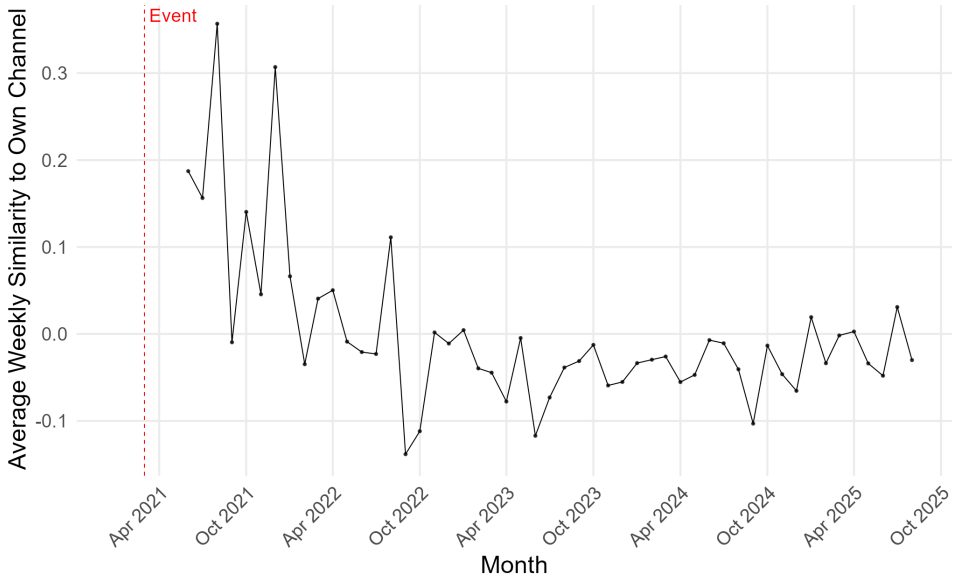
Table B.2: Effect of Shorts on Skipping to Chapter Markers

	Proportion Clicking	
	(1)	(2)
Chapter Slice	0.0171***	0.0271***
	(0.0010)	(0.0013)
Post × Chapter Slice	0.0530***	0.0573***
	(0.0015)	(0.0019)
Video Fixed Effects	Yes	Yes
Slice × Uploader Fixed Effects	Yes	Yes
Duration Restriction	No restriction	> 30 minutes
Observations	5,622,600	1,422,300

Notes: Table B.2 estimates the effect of the post-*Shorts* period on skipping activity to slices marked with a chapter, relative to unmarked slices. All specifications include video and slice × uploader fixed effects. Because whether a video falls before or after the introduction of *Shorts* does not vary within a video, the video fixed effects absorb the overall level effect of *Shorts* on skipping, so the reported coefficient captures only the additional, differential increase in skipping directed at chapter markers. This effect is positive and highly significant in both columns, and is robust to restricting the sample to videos over 30 minutes long in column (2). The effect of chapters on skipping is somewhat lower than, but of a similar magnitude to, the main slice-level skipping effects found throughout the paper, consistent with chapter markers being one further, particularly salient, entry point for content-seeking behaviour. Standard errors are clustered at the video level.

B.6 Supply Side and Substitution

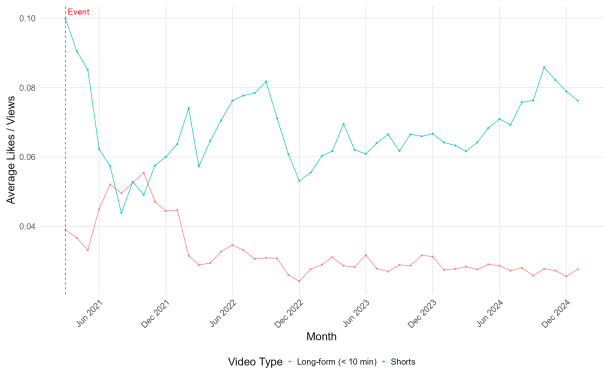
Figure B.29: Convergence of Shorts Content



Notes: Figure B.29 plots the difference between the cosine similarity from a *Short* to its parent channel and the cosine similarity from a *Short* to other *Shorts*, over time. As this difference shrinks towards zero, *Shorts* become as similar to each other as to their own parent channel, consistent with *Shorts* content becoming more formulaic over time. They are slightly negative, but the key insight is the trend rather than specific values.

Figure B.30: Engagement and Text Density, *Shorts* Versus Long-Form

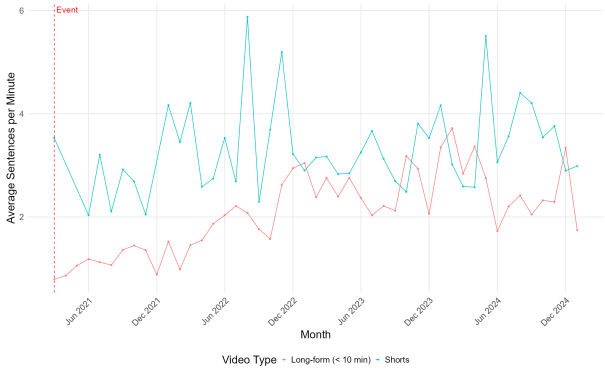
Panel A: Likes per View



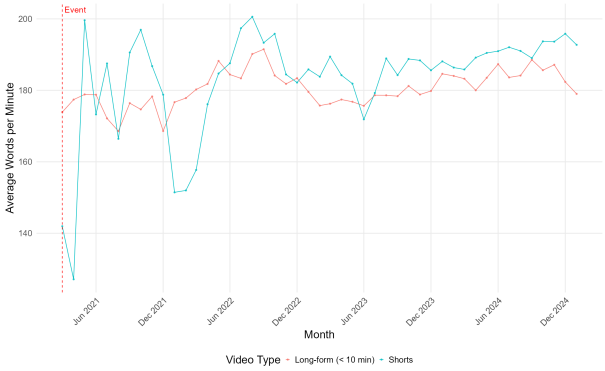
Panel B: Comments per View



Panel C: Sentences per Minute

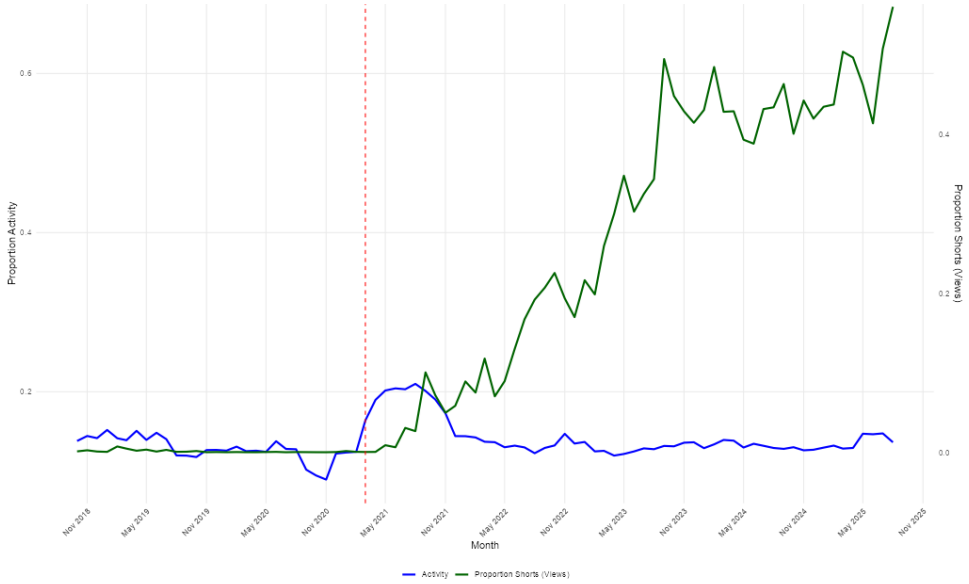


Panel D: Words per Minute



Notes: Figure B.30 plots the average of four outcomes over time, comparing *Shorts* to videos 10 minutes or less. This is done as these videos are the strongest substitutes given empirical results. Panel A and Panel B show that *Shorts* consistently get more likes per view but fewer comments per view. This suggests a change in the level of engagement with the news away from discussion. Additionally, Panel C and Panel D show that *Shorts* have more sentences and words per minute suggesting more information is compressed in a shorter time frame. These graphs are suggestive, but there is a sustained difference for these measures showing *Shorts* vary compared to the videos they replace (for some users).

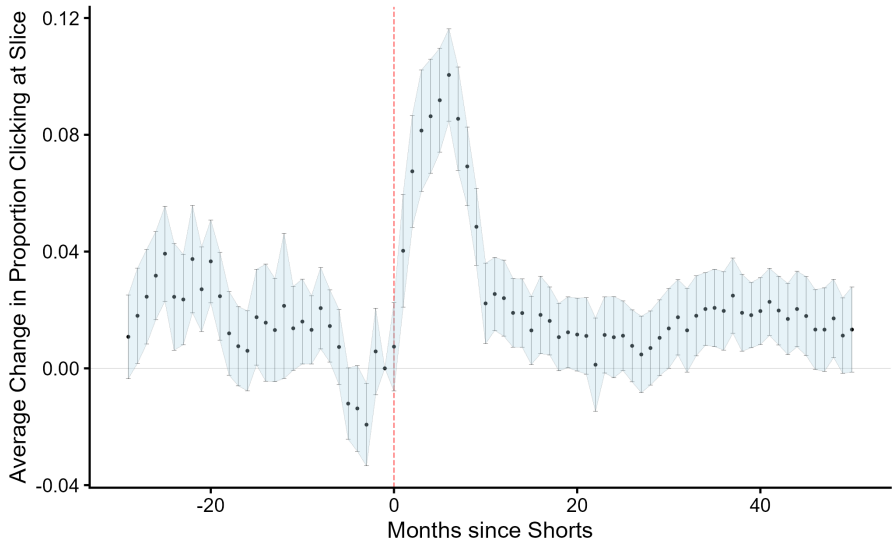
Figure B.31: Average Proportion of Views from Shorts vs Average Activity at Slice



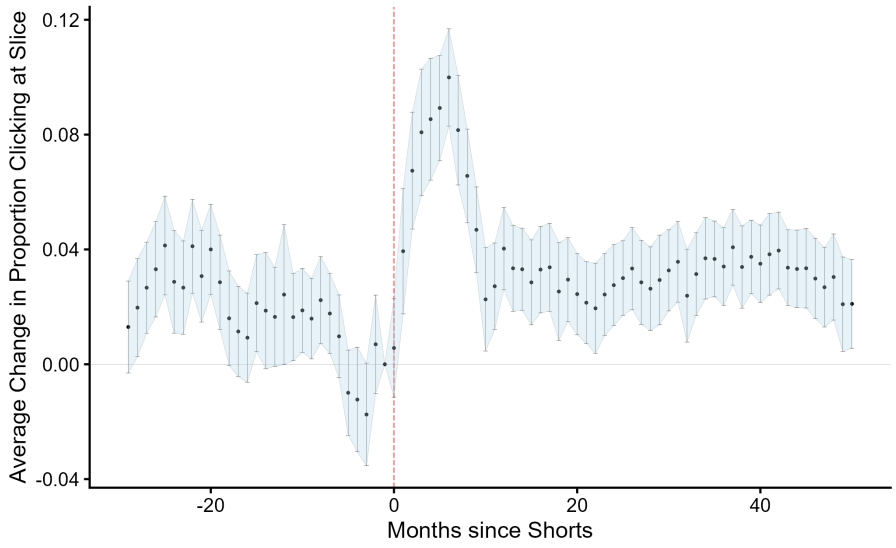
Notes: Figure B.31 takes the monthly average proportion of views received by *Shorts* relative to normal videos within channels in my sample, similar to Figure 16, and overlays the average activity at a slice, similar to Figure 4. If these two series were unrelated, they ought not to coincide. However, the initial spike in *Shorts* watching and the fall in activity for the shortest videos coincide almost exactly. This does not prove that they are related, but the empirical evidence is consistent with this interpretation. The alternative explanation, a sudden decay in the skipping effect unrelated to *Shorts* uptake, is possible but hard to reconcile with the sustained effect for longer videos. Overall, there is suggestive evidence that users substitute from skipping through short videos to *Shorts* when it becomes available.

Figure B.32: Supply Side: Controlling for Own Shorts Production

Panel A: Overall

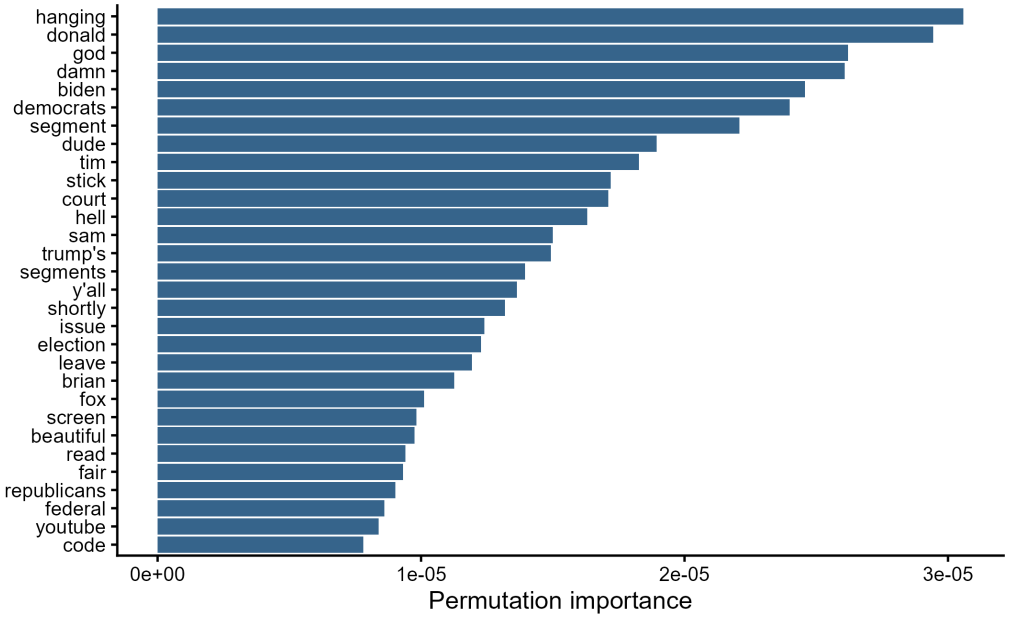


Panel B: Controlling for Proportion of Shorts



Notes: Figure B.32 estimates Equation 9 on the shortest videos. Panel A shows the overall effect, which decays over time. Panel B additionally controls for the proportion of *Shorts* produced by a channel in a given week, resulting in an overall consistent effect even for the shortest videos. It does not completely remove the decay, as I have access only to a subsample of available *Shorts* and it is likely a complex decision rule as to whether full switching occurs. Regardless, there is a higher sustained overall effect when controlling for the proportion of *Shorts* interacted with uploader: estimation partials out the outcome variable with respect to Proportion × Uploader, then re-estimates Equation 9 as usual. Confidence intervals are clustered at the video level, at 95% coverage.

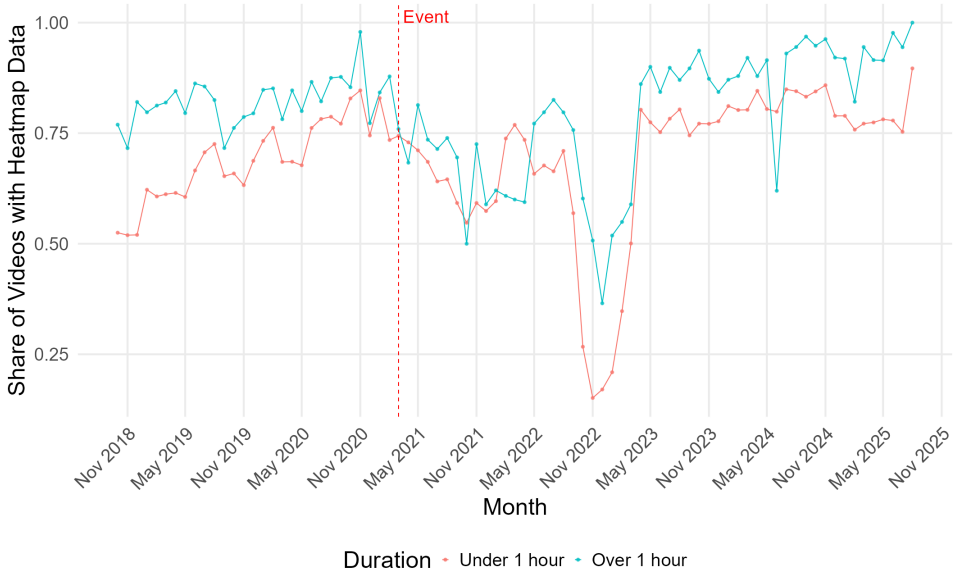
Figure B.33: Words Most Predictive of Skipping, by Permutation Importance



Notes: Figure B.33 is purely descriptive, with no fixed effects, no causal interpretation and differs to the DDML estimates in Table 6. A plain random forest is fit on the tf-idf matrix of the top 3,000 words across the full sample of videos to predict the proportion clicking at slice directly, and importance is measured by the fall in out-of-bag prediction accuracy when a word's values are permuted. There does not seem to be any words or overall topics that strongly suggest a potential for a content channel.

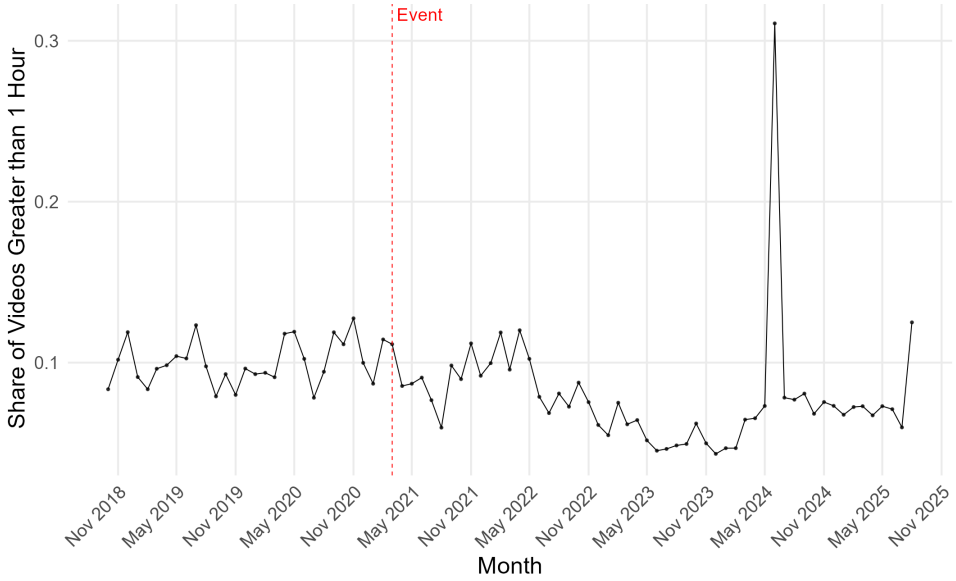
B.7 Alternative Explanations

Figure B.34: Probability Video had Heatmap Data



Notes: Figure B.34 shows the effect on the probability that a given video has heatmap data, to ensure that the main results are not being driven by a sudden change in selection of videos into having data. Whilst there is a clear break in Figure 5, there is no significant change here, ameliorating concerns of video selection into data availability. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.35: Probability Video is Greater than 1 Hour



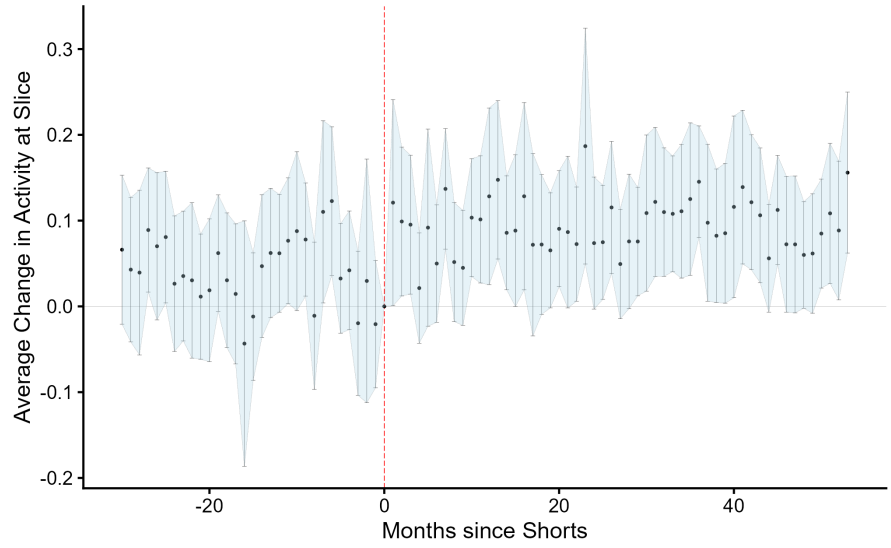
Notes: Figure B.35 shows the effect on the probability that a given video is treated (i.e. greater than 1 hour). This is to ensure that effects are not being driven by channels changing which videos are of certain lengths, or a general shift towards shorter or longer videos, around the policy. There is no jump around the policy, and results are robust to DDML on video transcript data, as shown in Table 6. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.36: Robustness to Sentence-Crossing Threshold

Panel A: Overall

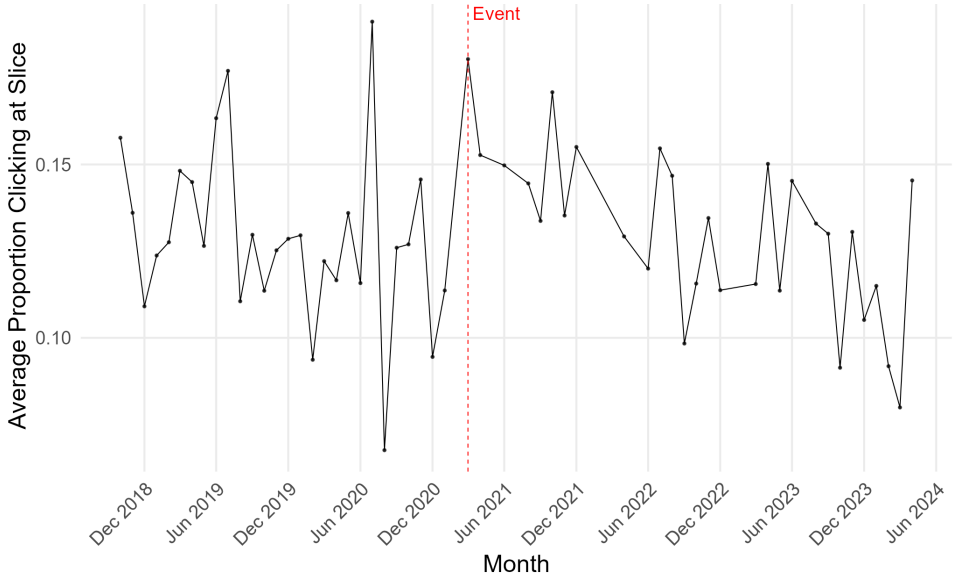


Panel B: Long Vs Short



Notes: Figure B.36 restricts to slices where at least 80% of sentences cross the slice threshold, ensuring the results reflect a change in context consumption rather than an artefact of the slice boundary. Panel A shows the overall effect, Panel B compares long and short videos; both are consistent with the baseline results, though with larger confidence intervals due to the smaller sample. Confidence intervals are clustered at the video level, at 95% coverage.

Figure B.37: MrBeast: Effect on Skipping



Notes: Figure B.37 shows the raw data of activity at a given slice for *MrBeast*'s videos around the introduction of *Shorts* for videos greater than 10 minutes. The effect is still there for all videos, but not as clear, lining up with the results in Figure 9. The plot suggests that skipping rises after the policy, but decays roughly linearly rather than showing the sharp drop-off seen for news channels, consistent with a more gradual re-adjustment towards attention-grabbing content for entertainment videos.

C Theoretical Extensions

This appendix extends the model in Section 3 in two directions. First, I characterise the second order conditions of the non-naive short consumption. The first allows skipping to feed back into short-form exposure. The second allows for skipped videos to additionally affect context value, and the third relaxes the assumption that users know the quality of a video.

C.1 Second Order Conditions

Starting from equation 6.

$$W(e) = f(e) + U(\phi(e)), \qquad U'(\phi) = C(\phi) \equiv \int_{S^*(\phi)} \int_{H_i(s)} \kappa_v(s'|s)\, ds'\, ds \tag{15}$$

As before. Rewrite the first derivative where $D(e)$ is the marginal damage of shorts exposure. This is a function of both how *Shorts* exposure affects the value of context and the value of context available. I could instead assume it was constant for simplicity here, but I allow it to vary:

$$W'(e) = \underbrace{f'(e)}_{\text{marginal gain}} - \underbrace{\left|\phi'(e)\right| C(\phi(e))}_{\equiv D(e),\ \text{marginal damage}} \tag{16}$$

Which gives the first order condition as before assuming an interior solution. The second derivative:

$$W''(e) = \underbrace{f''(e)}_{<0,\ \text{gain falls}} - \underbrace{D'(e)}_{\leq 0,\ \text{damage falls}} \tag{17}$$

$$D'(e) = -\Big[\underbrace{C'(\phi)\,\phi'(e)^2}_{\text{adaptation},\ \geq 0} + \underbrace{C(\phi)\,\phi''(e)}_{\text{damage decelerates},\ \geq 0\ \text{if}\ \phi\ \text{convex}}\Big] \tag{18}$$

Is not necessarily negative. The intuitive reason is two fold. First, adaption in response to *Shorts* exposure. If users now know that context value is harmed, they will watch different videos and parts of videos and if this readjustment lowers the harm by less than the gain of *Shorts*, users will consume more *Shorts*. The second reason is if the damage from *Shorts* decays relative to the gain from *Shorts*. Eventually, a corner solution where users only consume *Shorts* is possible. Therefore, to ensure a first order condition is indeed the global maximum we require:

$$\underbrace{\left|D'(e)\right|}_{\text{how fast damage falls}} < \underbrace{\left|f''(e)\right|}_{\text{how fast enjoyment falls}} \tag{19}$$

However, in both cases, there is still less *Shorts* consumption than in the naive regime, and hence the overall point still stands.

C.2 Dynamics

The previous results establish that naive users choose $e^{\text{naive}} > e^*$, increasing skipping through Proposition 1. A natural extension is allowing heavily skipped videos to fail to lower the overall exposure and improve focus. Let $n_v(S^*)$ denote the number of disjoint intervals (disconnected segments) in the optimal viewing set S^*. $n_v = 1$ corresponds to linear watching, and $n_v > 1$ reflects skipping. The equation updating e after watching video v is:

$$e' = e + \alpha s - \frac{\beta}{n_v(S^*)} \tag{20}$$

where αs reflects the increase in e from *Shorts* consumption s. If a user consumes more *Shorts*, their exposure increases. $\beta/n_v(S^*)$ reflects the restorative effect of watching normal video content, which is decreasing in how much the user skips. When $n_v = 1$, corresponding to linear watching, the full restorative effect β is realised. As $n_v \to \infty$, the video is consumed so fragmentarily that $\beta/n_v \to 0$ and no restoration occurs. The video is fragmented and keeps context value low.

Two conditions govern the long-run behaviour of e:

Corollary 1. *e is non-increasing after watching video v if and only if:*

$$n_v(S^*) \leq \frac{\beta}{\alpha s} \tag{21}$$

The user skips sufficiently little that the restorative effect of sequential viewing dominates short-form exposure. Otherwise, e drifts upward.

Corollary 2. *There exists a threshold $\bar{e}$ such that for $e > \bar{e}$, $\phi(e)$ is sufficiently low that $n_v(S^*)$ exceeds $\beta/\alpha s$. Above $\bar{e}$, e is self-sustaining: short-form exposure induces skipping, skipping prevents restoration, and e drifts upward indefinitely. Below $\bar{e}$, users who watch sufficiently consecutive long-form content can restore e toward zero.*

The introduction of *YouTube Shorts* provides the initial shock that increases overall skipping behaviour through a direct effect on the value of context in news media. In addition, if we believe that long-form content consumed like short-form content may have a similar effect, the dynamic is self-sustaining. Short-form exposure erodes context formation, which induces more skipping through normal videos, which prevents restoration, which sustains high e.

I do not suggest that this is sustained across browsing sessions. Rather, sufficient *Shorts* exposure may leave users trapped in a content seeking loop even when watching longer content for the remainder of their given watching session. In equilibrium, users continually expose themselves to *Shorts* and the effect is sustained.

C.3 Learning about Quality

The model in Section 3 assumes users know the utility of every slice. Here I relax this and allow users to learn about the quality of a video as they watch it. Learning amplifies the effect as when context is worth less, learning about it is worth less. For this extension it is convenient to work with discrete slices $s = 1, \ldots, N$, matching the data.[77]

Quality Each video has a content quality θ_v^q and a context quality θ_v^c, drawn independently from

$$\theta_v^q \sim N(\mu_q, \sigma_q^2), \qquad \theta_v^c \sim N(\mu_c, \sigma_c^2) \tag{22}$$

Users know these distributions but not the draw for any given video. Content and link values vary within a video around its quality:

$$\underbrace{q_v(s)}_{\text{content of slice } s} = \theta_v^q + \varepsilon_v(s), \qquad \underbrace{\kappa_v(s'|s)}_{\text{link between } s' \text{ and } s} = \theta_v^c + \eta_v(s', s) \tag{23}$$

where $\varepsilon_v(s) \sim N(0, \tau_q^2)$ and $\eta_v(s', s) \sim N(0, \tau_c^2)$ are independent. I abstract from learning across videos. Users watch only a small subset of all videos, so a run of bad videos tells them little about the population, and their prior resets to (μ, σ^2) for each new video. Additionally, the context link is simply a fixed average for ease in order to show the trade-offs.

Learning Within a Video Users update by Bayes' rule as they watch. Importantly, they observe the content and context quality of a slice separately. This could be relaxed but is a useful simplifying assumption. Write $\mathbb{E}_S[\cdot]$ for the expectation given the set of slices S watched so far, which contains n slices. Expected content quality is

$$\underbrace{\hat{\theta}_v^q(S) \equiv \mathbb{E}_S\left[\theta_v^q\right]}_{\text{expected content quality}} = \underbrace{(1-\omega_q)\,\mu_q}_{\text{prior}} + \underbrace{\omega_q\,\bar{q}_v(S)}_{\text{average content watched}}, \qquad \omega_q = \frac{n/\tau_q^2}{1/\sigma_q^2 + n/\tau_q^2} \tag{24}$$

Expected content quality is a weighted average of the prior mean and the average content watched so far. As more slices are watched, more weight is placed on observed video content. Note that except for n, all other parameters are exogenous.

Context quality is learnt in the same way, except that users observe links rather than slices. Expected context quality $\hat{\theta}_v^c(S) \equiv \mathbb{E}_S\left[\theta_v^c\right]$ is a weighted average of μ_c and the average link value observed, with weight $\omega_c = \frac{m/\tau_c^2}{1/\sigma_c^2 + m/\tau_c^2}$, where m is the number of links observed.

[77] The goal is to reflect the key trade-off when quality is unknown.

Skip Decision As before, the decision to skip reflects a lower value of context relative to content. These values are taken in expectation, but nonetheless the same trade-off holds.

Proposition 2. *A user at slice s_0 having watched S jumps to slice s whenever*

$$\underbrace{\mathbb{E}_S\big[q_v(s) - q_v(s_0^+)\big]}_{\textit{expected content gain}} > \underbrace{\mathbb{E}_S\big[c(s_0^+ \mid H) - c(s \mid H)\big]}_{\textit{expected context lost}} \tag{25}$$

The expected content gain is a difference in $\hat{\theta}_v^q(S)$ and the expected context lost is $\hat{\theta}_v^c(S)$ times the links broken by jumping, so the trade-off is that of Section 3 with expected qualities in place of known ones.

Skipping and the Speed of Learning n and m differentially respond to skipping. Every slice watched adds one observation of content. A slice only adds observations of context through its links to the other slices watched, so

$$\underbrace{n = |S|}_{\text{content observations}} \quad , \quad \underbrace{m = \frac{n(n-1)}{2}}_{\text{context observations}} \tag{26}$$

Proposition 3. *Consider a stretch of N slices that a user either watches in full or skips through, watching a fraction x of them. Relative to watching in full, skipping reduces the information gained about content quality by a factor of x, but the information gained about context quality by a factor of approximately x^2.*

With $n = xN$ the share of links observed is $\frac{xN(xN-1)}{N(N-1)}$, which is x^2 up to the -1. Skipping half of a video halves what a user learns about its content but quarters what they learn about its context. A user who skips therefore learns mostly about content. Both ω_q and ω_c fall when a user skips, but ω_c falls faster, so expected context quality lags behind expected content quality. If link values also decline with the distance between slices, as seems natural for a narrative, the asymmetry is stronger still since skipping removes the nearest and most valuable links. Note that if the pairwise assumption were relaxed, and allowed for triplets and quadruplets, the context learning would be even smaller.

Shorts and the Value of Learning Before *Shorts*, a user deciding how to watch faces a trade-off. Skipping lets them reach the content they want, but sacrifices learning about context. After *Shorts*, context is worth less, so learning about it is worth less, and giving it up by skipping costs less. More users become content seekers. This operates on top of the direct effect in Section 3, where a lower $\phi(e)$ lowers the value of context itself.

Learning Trap Example To see how skipping could sustain itself, consider a simple example. A user has the choice over a stretch of video between watching consecutively and skipping. I abstract from finishing the video sooner, and simply consider the trade-off between watching in full or searching for content. The expected context from watching consecutively is $c(\hat{\theta}_v^c \mid S)$, the context function of Section 3, multiplied by $\phi(e)$ as usual and increasing in expected context quality. There is therefore a cutoff $\bar{c} > 0$, which depends on the content forgone, such that the user watches consecutively if and only if

$$\phi(e)\, c\big(\hat{\theta}_v^c \mid S\big) \geq \bar{c} \quad \Longleftrightarrow \quad \hat{\theta}_v^c \geq t(e) \tag{27}$$

where $t(e)$ is the expected context quality at which the user is indifferent. Since $\phi'(e) < 0$, *Shorts* raises the cutoff $t(e)$.

On average the links a user observes have value θ_v^c, so in a video of true context quality θ_v^c expected context quality after m links is $\mathbb{E}\big[\hat{\theta}_v^c \mid \theta_v^c, m\big] = (1-\omega_c)\,\mu_c + \omega_c\,\theta_v^c$. A user who is currently skipping switches to watching consecutively only once this reaches $t(e)$, which requires

$$\theta_v^c \geq \bar{\theta}(m,e) \equiv \mu_c + \big(t(e) - \mu_c\big)\left(1 + \frac{\tau_c^2}{\sigma_c^2\, m}\right) \tag{28}$$

When $t(e) > \mu_c$, $\bar{\theta}(m,e)$ lies strictly above $t(e)$, and the gap closes only as m grows.

Corollary 3. *Suppose $\mu_c < t(e)$, so that a user begins a video by skipping. The user expects to continue to skip in any video with*

$$\underbrace{t(e) \leq \theta_v^c}_{\text{worth watching consecutively}} \qquad < \qquad \underbrace{\bar{\theta}(m,e)}_{\text{not yet discovered}} \tag{29}$$

This region has width $\big(t(e) - \mu_c\big)\,\tau_c^2/(\sigma_c^2\, m)$, which is increasing in e and decreasing in the number of links observed m. To escape it, a user must observe at least

$$m^*(\theta_v^c, e) = \frac{\tau_c^2}{\sigma_c^2} \cdot \frac{t(e) - \mu_c}{\theta_v^c - t(e)} \tag{30}$$

links, which is also increasing in e. Since m falls with the square of the fraction of slices watched, skippers are the users least likely to reach m^*.

Before and After Shorts Suppose that before *Shorts*, when $\phi(0) = 1$, the average video is worth watching consecutively, so $\mu_c \geq t(0)$. Users begin by watching consecutively, observe links quickly, and learn θ_v^c. They end up skipping only when $\theta_v^c < t(0)$, and no trap exists. Now suppose exposure raises the cutoff above the prior, $t(e) > \mu_c$. Users

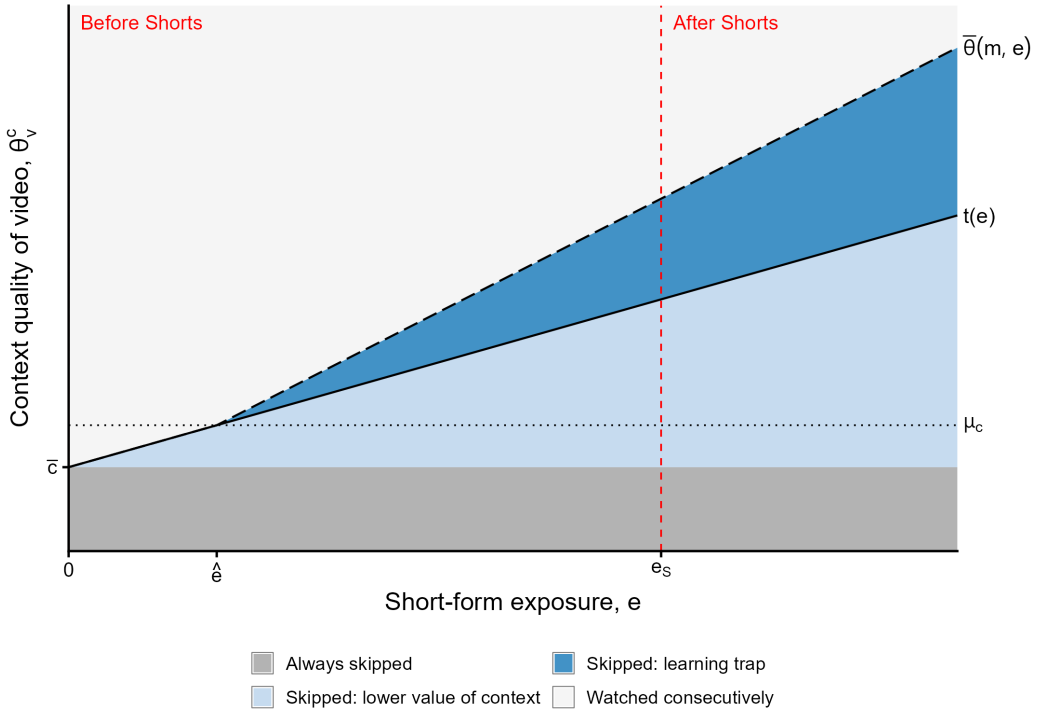
begin by skipping, and videos fall into four groups according to their context quality:

$$\begin{aligned}
&\theta_v^c < t(0) && \text{skipped before and after } \textit{Shorts} \\
&t(0) \leq \theta_v^c < t(e) && \text{skipped due to lower } \phi(e) \\
&t(e) \leq \theta_v^c < \bar{\theta}(m,e) && \text{skipped due to learning} \\
&\theta_v^c \geq \bar{\theta}(m,e) && \text{watched consecutively}
\end{aligned} \tag{31}$$

The second group reflects the direct effect in Section 3, where context is genuinely worth less. The third group is the learning channel. These videos would be watched consecutively by a user who knew their quality, but are skipped because the user never observes enough links to find out. This is more likely after *Shorts* because the lower value of context makes learning about context more costly. Once a user starts skipping, it is sustained for exactly these videos. Figure C.1 plots these regions against exposure.

Implications Learning amplifies the effect of *Shorts*. It predicts an increase in skipping larger than the fall in the value of context alone would imply. Since the information lost by skipping grows with the square of the slices watched, it also predicts that longer videos, which have the most context to learn about, are the most affected, consistent with Figure 8.

Figure C.1: Skipping Regions With Learning, Before and After *Shorts*



Notes: Figure C.1 plots the regions in Equation 31 against short-form exposure e. This abstracts from the amount of learning, and asks whether the video is skipped or watched consecutively in the end. The solid line is the full-information cut-off $t(e) = \bar{c}/\phi(e)$ and the dashed line is the discovery threshold $\bar{\theta}(m, e)$ from Equation 28. The learning trap opens at $\hat{e}$, where $t(\hat{e}) = \mu_c$. Before *Shorts* only videos with $\theta_v^c < \bar{c}$ are skipped. After *Shorts*, at exposure e_S, videos are additionally skipped because context is worth less (light blue) and because users never learn that a video's context is worth watching for (dark blue). The functional form $\phi(e) = 1/(1+e)$ and all parameter values are illustrative.

C.4 Ordering and Switching Costs

This extension considers ways to improve consumption utility without changing content, only its ordering. We assume as in Section 3 that users are naive and do not internalise the context cost of short-form videos. A session consists of an amount e of *Shorts* and long videos. For simplicity, I will assume just one long video to make the trade-off clearer, but the logic holds more generally. The utility of consuming long videos is written only as a function of ϕ, the usual context cost from exposure. *Shorts* are pure content, so their value does not depend on ϕ.

A naive user either consumes greedily, taking whichever option currently offers the higher utility, or randomly. This likely results in users consuming some short and some long videos in succession, so that videos are watched at some average capacity $\bar{\phi}$, with $\phi(e) < \bar{\phi} < 1$:

$$\underbrace{W_I}_{\text{interleaved session}} = \underbrace{f(e)}_{\textit{Shorts}} + \underbrace{U(\bar{\phi})}_{\text{video at average capacity}} \tag{32}$$

Compare this to consuming all long videos first and all *Shorts* afterwards:

$$\underbrace{W_L}_{\text{video first}} = \underbrace{f(e)}_{\textit{Shorts}} + \underbrace{U(1)}_{\text{video at full capacity}} \tag{33}$$

The same quantity of *Shorts* and the same videos, reordered, give

$$W_L - W_I = \underbrace{U(1) - U(\bar{\phi})}_{\text{context saved by reordering}} > 0 \tag{34}$$

The gain comes entirely from watching the videos before capacity has been degraded. The difficulty is that a naive user will not reorder on their own, as doing so requires internalising ϕ'.[78]

To potentially ameliorate this issue, we might introduce a cost c of switching between the two types. A naive user would not internalise the cost of short-form content exposure, but if they are forward looking enough to plan their consumption they immediately would choose *Shorts* as a single block at one end of the session, either *Shorts* first (S) or the video first (L):

$$\underbrace{W_S}_{\textit{Shorts}\text{ first}} = f(e) + \underbrace{U\big(\phi(e)\big)}_{\text{video at degraded capacity}} - \underbrace{c}_{\text{one switch}}, \qquad \underbrace{W_L}_{\text{video first}} = f(e) + \underbrace{U(1)}_{\text{video at full capacity}} - \underbrace{c}_{\text{one switch}} \tag{35}$$

[78] Additionally, if *Shorts* are addictive and start with high utility, a user may struggle to watch long videos first even if in the long run this is better for them.

Both regimes involve exactly one switch instead of two hence their optimality.

$$W_L - W_S = \underbrace{U(1) - U\big(\phi(e)\big)}_{\text{context lost by watching } \textit{Shorts} \text{ first}} > 0 \tag{36}$$

Here, we need to assume that if given the choice between the two regimes, users would be able to realise that consuming *Shorts* last is better. Perhaps it is harder to notice the effect of one additional short, but clearer between these two regimes. This is improving if:

$$\underbrace{U(1) - c}_{\text{video first, bunched}} > \underbrace{U(\bar{\phi})}_{\text{interleaved}} \iff c < U(1) - U(\bar{\phi}) \tag{37}$$

Alternatively, if users genuinely cannot differentiate between the two regimes, then assuming a random choice at a tie we need:

$$\underbrace{\tfrac{1}{2}\big(W_L + W_S\big)}_{\text{random regime}} > \underbrace{W_I}_{\text{interleaved}} \iff c < \tfrac{1}{2}\big[U(1) + U(\phi(e))\big] - U(\bar{\phi}) \tag{38}$$

This is a tighter requirement, as the user now only lands in the better regime half of the time.

Note that this is potentially not feasible in such an extreme case. It requires users who may be over consuming *Shorts* to plan their consumption well in advance. However, if switching is harder through fewer recommendations of *Shorts* from long videos and vice versa, it could at least somewhat reduce the harm to context valuation. Whether that aligns with platform interests or could even be incentive compatible given the potential for addiction is of course a further question beyond the scope of this paper.

D Structural Model

The first part maps the observed shift in skipping to the value of context and states the assumptions it requires. The second estimates which destinations become relatively more attractive as that value falls. The third sets out a discrete-choice specification of the change in the relative value of context, though requires stronger assumptions hence is not the main specification.

D.1 Mapping Skips to Context Value

D.1.1 Setup

Time is indexed by $t \in \{0, 1\}$, the pre- and post-*Shorts* periods, with $\phi_0 \equiv \phi(0) = 1$ and $\phi_1 \equiv \phi(e) < 1$. By Equations 1 and 5, the utility of slice s in period t, given history H_i, is

$$u_t(s \mid H_i) = q_v(s) + \phi_t \, C_v(s \mid H_i), \qquad C_v(s \mid H_i) \equiv \int_{H_i(s)} \kappa_v(s'|s) \, ds'. \tag{39}$$

At a *skip opportunity*, a viewer who has watched up to s_0 decides whether to continue to s_0^+ or jump to a destination s. The two options differ, following (39) evaluated at each, by

$$u_t(s \mid H_i) - u_t(s_0^+ \mid H_i) = \underbrace{\left[q_v(s) - q_v(s_0^+)\right]}_{\equiv \, \Delta q_{sv}} - \phi_t \underbrace{\left[C_v(s_0^+ \mid H_i) - C_v(s \mid H_i)\right]}_{\equiv \, D_s}. \tag{40}$$

Δq_{sv} is the content gained by jumping and D_s is the context given up. Continuing along the path builds context, while jumping elsewhere may not, so D_s is typically positive (made exact below). The viewer skips if (40) is positive:

$$\Delta q_{sv} > \phi_t \, D_s, \tag{41}$$

which is Equation 12 with $\Delta c_s \equiv \phi_t \, D_s$. Activity in period t, a_t, is the population share of opportunities satisfying (41), aggregated across viewers, videos and positions:

$$a_t = \Pr\left(\Delta q_{sv} > \phi_t \, D_s\right). \tag{42}$$

Each viewer is exposed to different values, hence skipping is not necessarily 0 or 1 only.

D.1.2 Assumptions

Three assumptions are required.

Non-trivial stakes $\Delta q_{sv} > 0$ and $D_s > 0$ for every opportunity. Every destination offers some content, and every skip gives up some context link. If this does not hold, then in truth the increase of 0.08 reflects a far greater change in the relative value of content and context, but this assumption makes the poles (0 and 1) the genuine lower and upper bounds.

Content unchanged and context-stock stability $\mathrm{c}\Delta q_{sv}$ and D_s are time-invariant, and their joint distribution across opportunities is stable across periods. This seems reasonable given no evidence for re-adjustment by channels and an immediate effect.

Independent exponential stakes $\Delta q_{sv} \sim \mathrm{Exp}(\mu_q)$ and $D_s \sim \mathrm{Exp}(\mu_D)$, independently across and within opportunities.

D.1.3 Main Result

Lemma 1 (Poles)**.** *Under Non-trivial stakes,* $a_t \to 1$ *as* $\phi_t \to 0$ *and* $a_t \to 0$ *as* $\phi_t \to \infty$*, for any distribution of* $(\Delta q_{sv}, D_s)$*.*

Proof. From (42), $a_t = \Pr(\Delta q_{sv}/D_s > \phi_t)$. As $\phi_t \to 0$ the event holds almost surely since $\Delta q_{sv}/D_s > 0$; as $\phi_t \to \infty$ it fails almost surely since $\Delta q_{sv}/D_s$ is finite. □

Proposition 4 (Logistic skip share)**.** *Under Independent exponential stakes,*

$$a_t = \frac{1}{1 + \kappa\,\phi_t}, \qquad \kappa \equiv \mu_D/\mu_q, \tag{43}$$

so a_t *is a logistic function of* $\ln \phi_t$ *with slope exactly* 1*:* $\ln\big(a_t/(1-a_t)\big) = \ln(\mu_q/\mu_D) - \ln \phi_t$*.*

Proof. Condition on D_s. Since $\Delta q_{sv} \sim \mathrm{Exp}(\mu_q)$, $\Pr(\Delta q_{sv} > x) = e^{-x/\mu_q}$, so

$$a_t = \mathbb{E}_{D_s}\left[e^{-\phi_t D_s/\mu_q}\right] = \int_0^\infty \frac{1}{\mu_D} e^{-d/\mu_D}\, e^{-\phi_t d/\mu_q}\, dd = \frac{1}{1 + \phi_t\,\mu_D/\mu_q}. \tag{44}$$

□

D.1.4 Interpretation Relative to the Pre-Period

Corollary 4 (Identification of ϕ_1)**.** *Normalising* $\phi_0 = 1$ *at the pre-period mean* a_0 *pins* $\kappa = (1 - a_0)/a_0$ *in* (43)*, and identifies*

$$\phi_1 = \frac{a_0/(1 - a_0)}{a_1/(1 - a_1)}, \tag{45}$$

the ratio of pre- to post-period skipping odds, where a_1 *is the post-period mean. The value of context relative to content falls from* 1 *to* ϕ_1*, a fall of* $1 - \phi_1$*. Equivalently, content rises in value relative to context by* $1/\phi_1 - 1$*.*

Proof. Immediate from (43) evaluated at $t = 0, 1$ and dividing. □

At $a_0 = 0.15$ and a post-period mean $a_1 = 0.23$ (the 0.08 mean shift of Figure 5), $\phi_1 \approx 0.59$: the value of context falls by about 41%.

D.1.5 Aleternative Reference

Corollary 4 indexes the change relative to the pre-period mean level of skipping. This alternative specification takes the approach of assuming a lower bound for the index.

Context is evaluated continuously, whereas content is evaluated at a single slice only. Therefore, there is an asymmetry in outcomes even if content and context were equally valued. At the extreme, with content and context equally valued, the highest skipping could be is 0.5. This would require context to never aggregate more than one slice previous, which is an extreme assumption. However, given that context aggregates weakly, the index cannot possibly be greater than 0.5.

Applying this to the same 0.08 shift, read instead as $a : 0.5 \to 0.58$: $\phi_1 \approx 0.72$, a fall in the value of context of at least 28%. Context lost at least 28% of its value, and about 41% under the pre-period-anchored reading.

D.2 Modelling Destinations

Chapter headings are a useful benchmark destination. They likely carry far less context relative to what preceded them, so whatever a viewer gains by landing there is mostly content. A more extreme version is that context completely resets at a new chapter heading. This is probably true for videos that are a collection of independent news stories, but not likely to hold universally. The following stylised model shows that in general the fall in the value of context benefits chapters disproportionately.

Setup Let c denote a chapter heading and s a slice at depth τ_s since the most recent heading, both in the same video. Context completely resets[79] at each heading and builds only within a chapter, so $C_v(c \mid H_i) = 0$ and $C_v(s \mid H_i) = G(\tau_s)$ for some $G : \mathbb{R}_+ \to \mathbb{R}_+$ with $G(0) = 0$, non-decreasing. Landing utilities are then

$$u_{ct} = q_v(c), \qquad u_{st} = q_v(s) + \phi_t\, G(\tau_s). \tag{46}$$

Proposition 5 (Chapters as the maximal-relative destination)**.** *Under Content Unchanged, the advantage of a heading over a slice at depth τ_s,*

$$u_{ct} - u_{st} = \big(q_v(c) - q_v(s)\big) - \phi_t\, G(\tau_s), \tag{47}$$

[79] This can be relaxed; all that is required is relatively fewer context links at new chapters. However, in the extreme case context is identified exactly.

is non-increasing in ϕ_t, strictly so whenever $\tau_s > 0$. A heading's relative advantage over every other destination in its chapter is therefore largest exactly where it can be, at the heading itself, where it is zero by definition, and it only grows as ϕ_t falls.

Proof. $\partial(u_{ct} - u_{st})/\partial\phi_t = -G(\tau_s) \leq 0$, with equality only at $\tau_s = 0$ (where $s = c$ and the gap is identically zero). Since $\phi_1 < \phi_0$, the gap is weakly larger in period 1 than in period 0, strictly so for $\tau_s > 0$. □

This needs no choice probability or distributional assumption on (q_v, G). Conditional on skipping, chapter headings are relatively more attractive destinations after *Shorts* than any point into a chapter. As context builds up, each new part is less relatively valuable as it contains content and context. This gives the shape as shown in Figure 14.

The shape of context If we take these assumptions seriously, then $G(\cdot)$ can be identified from the shock to the value of context. Take the same comparison across periods. Content Unchanged leaves $q_v(c) - q_v(s)$ fixed, so

$$\big(u_{c1} - u_{s1}\big) - \big(u_{c0} - u_{s0}\big) = (1 - \phi_1)\, G(\tau_s), \tag{48}$$

using $\phi_0 = 1$. The left-hand side is exactly what Figure 14 estimates at each depth bin and $(1 - \phi_1)$ is a constant shifter (already pinned down by Corollary 4). Reading the coefficient at depth τ as the left-hand side,

$$G(\tau) \;\approx\; \frac{\text{coefficient at } \tau}{1 - \phi_1}, \tag{49}$$

turns the depth profile into an estimate of the context a chapter has built up by τ, in the same units as content utility.

I non-parametrically estimate the functional form with bins of 30 seconds. Figure 14 flattens out after a few minutes. Whether that flattening reflects a true ceiling on narrative dependence or simply diminishing returns to depth is not something this exercise can distinguish, but the overall shape informs us of how users expect context to build up in online news videos. At the very least, it reveals which parts of the video benefit relatively more, which lines up exactly with the model.

D.3 An Alternative Discrete-Choice Specification

I alternatively treat the choice of destination *within* a video as a discrete choice among slices, in the style of Berry et al. (1995), Nevo (2001), and Train (2009), and identify a similar underlying object. As a baseline this estimates, within a video v,

$$\underbrace{\tilde{u}_{isv}}_{\text{Utility of Skipping}} = \underbrace{\lambda_i Shorts_{s<96} \times Post_t}_{\text{Effect of Shorts}} + \underbrace{f_{vs}}_{\text{Video/Slice FEs}} + \underbrace{\varepsilon_{isv}}_{\text{Type 1 EV}}, \tag{50}$$

where $Shorts_{s<96}$ is an indicator equal to one if the given slice is in the first 95% of the video leveraging the relative preference for content seeking after *Shorts*

D.3.1 Setup

Time is indexed by $t \in \{0, 1\}$, corresponding to the pre- and post-*Shorts* periods. Exposure satisfies $e_0 = 0$ and $e_1 = e > 0$, so that $\phi_0 \equiv \phi(0) = 1$ and $\phi_1 \equiv \phi(e) < 1$. At each decision node within video v, user i with history H_i chooses a destination slice from a choice set $\mathcal{J}_v \subseteq [0, 1]$ (which includes continuation and, implicitly, exit). Define the *context stock* of slice s as

$$C_v(s \mid H_i) \equiv \int_{H_i} \kappa_v(s'|s)\, ds', \tag{51}$$

so that the conditional indirect utility of choosing slice s in period t is

$$u_{isvt} = q_v(s) + \phi_t\, C_v(s \mid H_i) + \varepsilon_{isvt}, \tag{52}$$

where ε_{isvt} is i.i.d. Type 1 Extreme Value with scale σ_t. Note that the skipping utility in Equation 12 is the difference of (52) across two alternatives, so the two formulations are equivalent. Let $\bar{C}_{vst} \equiv \mathbb{E}\left[C_v(s \mid H_i) \mid v, s, t\right]$ denote the average context stock at slice s under the period-t equilibrium distribution of histories. Let $B \equiv \{s : s < 0.96\}$ denote the treated ("body") slices and $T \equiv \{s : s \geq 0.96\}$ the omitted ("tail") slices, matching the construction of $Shorts_{s<96}$.

D.3.2 Assumptions

Four assumptions are required for the main result.

Content Unchanged $q_v(s)$ is time-invariant, and *Shorts* exposure enters utility only through ϕ_t multiplying the context stock. In particular, there is no supply-side response (the content and composition of videos is unchanged) and exposure does not affect content utility directly. This is the formal statement of the argument in Section 5.2.1 that average video content does not change around the introduction of *Shorts*.

Context-stock stability $\bar{C}_{vst} = \bar{C}_{vs}$ for $t = 0, 1$: the average context stock at each video-slice, evaluated at equilibrium histories, is time-invariant. This requires that the context stocks at which choices are evaluated do not themselves adjust.

Scale invariance $\sigma_0 = \sigma_1$, normalised to 1: exposure does not change the dispersion of the idiosyncratic taste shocks. This rules out *Shorts* making choice behaviour uniformly noisier.

Composition The distribution of users and videos is stable across periods.

D.3.3 Main Result

Lemma 2. *Under (52) and Type 1 EV errors with scale σ_t, for any two slices s, s' available at the same decision node,*

$$\ln \frac{P_{vst}}{P_{vs't}} = \frac{1}{\sigma_t}\Big[\big(q_v(s) - q_v(s')\big) + \phi_t\big(\bar{C}_{vst} - \bar{C}_{vs't}\big)\Big]. \tag{53}$$

Proof. Immediate from the closed form of conditional logit choice probabilities, taking expectations over histories within the video-slice-period cell. □

Proposition 6 (Identification of λ)**.** *Under the four assumptions above, the coefficient on $Shorts_{s<96} \times Post_t$ in Equation 50, with video-slice fixed effects f_{vs} and period effects, identifies*

$$\lambda = \big(1 - \phi_1\big) \cdot \bar{D}, \qquad \bar{D} \equiv \bar{C}_T - \bar{C}_B, \tag{54}$$

where $\bar{C}_T$ and $\bar{C}_B$ are (estimation-weighted) averages of $\bar{C}_{vs}$ over tail and body slices respectively. λ is the proportional decline in the marginal value of context, scaled by the average context-stock gap between omitted and treated slices.

Proof. By Lemma 2 with $\sigma_t = 1$ (scale invariance), the log-odds of slice $s \in B$ relative to $s' \in T$ in period t are $\big(q_v(s) - q_v(s')\big) + \phi_t\big(\bar{C}_{vs} - \bar{C}_{vs'}\big)$, where context-stock stability drops the t subscript on context stocks and exclusion drops it on q. The difference-in-differences across periods is therefore

$$\big(\phi_1 - \phi_0\big)\big(\bar{C}_{vs} - \bar{C}_{vs'}\big) = \big(1 - \phi_1\big)\big(\bar{C}_{vs'} - \bar{C}_{vs}\big), \tag{55}$$

since $\phi_0 = 1$. The fixed effects f_{vs} absorb the time-invariant level $q_v(s) + \bar{C}_{vs}$, and the period effect absorbs shifts common to all slices. The interaction coefficient is then the weighted average of the slice-pair difference-in-differences, which is (54), with the composition assumption ensuring the weights are stable across periods. □

D.3.4 Interpretation as Relative Valuation

Corollary 5 (Content relative to context)**.** *The marginal utility of a unit of content is 1 in both periods, while the marginal utility of a unit of context stock is ϕ_t. The value of content relative to context therefore rises from 1 to $1/\phi_1$, an increase of*

$$\frac{1 - \phi_1}{\phi_1} = \frac{\lambda/\bar{D}}{1 - \lambda/\bar{D}}. \tag{56}$$

Under the normalisation $\bar{D} = 1$ (the context gap between tail and body is one unit, measured in units of content utility), this equals $\lambda/(1 - \lambda) \approx \lambda$ to first order.

Thus $\lambda/(1-\lambda)$ is the change in the value of content to context, *given* the unit normalisation $\bar{D} = 1$. Without a normalisation or an external estimate of $\bar{D}$ (e.g. from a parametrised κ_v fitted to pre-period linear-viewing behaviour), λ identifies only the product $(1-\phi_1)\bar{D}$. The level of ϕ_1 and the context gap are not separately identified from the single difference-in-differences.

D.3.5 Full Results

Table D.1 reports the resulting estimates of λ. Under Corollary 5's normalisation, content is about 15% more valuable than context, the same order of magnitude as the single-margin estimate above.

Table D.1: Structural Model of Skipping

	Log Skipping Share		
	(1)	(2)	(3)
λ	0.6174***	0.1317***	0.1309***
	(0.0107)	(0.0183)	(0.0222)
Video Fixed Effects	Yes	Yes	Yes
Upload Date Fixed Effects	Yes	Yes	Yes
Uploader × Slice Fixed Effects	No	Yes	Yes
Duration Restriction	> 30 minutes	>30 minutes	>60 minutes
Observations	1,331,822	1,331,822	690,608

Notes: Table D.1 estimates Equation 50, the effect of the post-*Shorts* period on log skipping share (following the discrete choice literature). All specifications include video and upload date fixed effects. Columns (2) and (3) additionally include uploader × slice fixed effects. The sample is restricted to videos with *duration* > 1800 seconds in columns (1) and (2) and *duration* > 3600 seconds in column (3). The coefficient remains positive and highly significant throughout, though it falls substantially once uploader × slice fixed effects are added. Given small values in skipping for some slices and larger for others, this likely controls for skipping to certain slices in general being common, which when normalising shares has a fundamental impact on the coefficient. Standard errors are clustered at the video level and reported in parentheses.