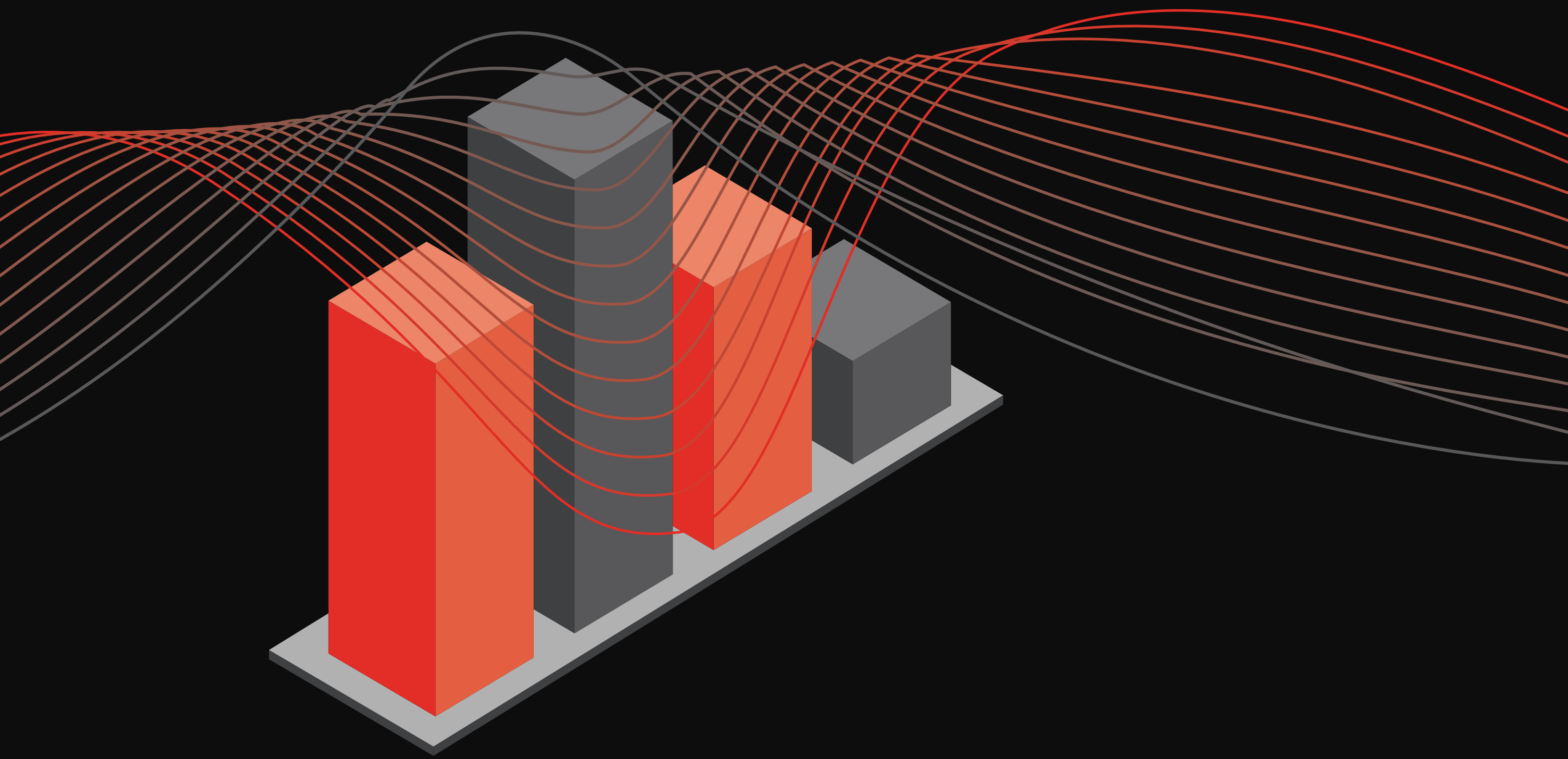
TECHNOLOGY POLICY INSTITUTE



WHERE DOES THE TIME GO? COMPETING FOR ATTENTION IN THE ONLINE ECONOMY

Scott Wallsten, Sarah Oh Lam, & Nathaniel Lovin

May 20, 2023

WHERE DOES THE TIME GO?
COMPETING FOR ATTENTION IN THE ONLINE ECONOMY

Scott Wallsten ‡ Sarah Oh Lam* Nathaniel Lovin§

May 20, 2022

Many of the goods and services available online are free in the sense that they do not require financial payments. But they do require investments of time, which is not free. Thus, the price of any time spent on a site online is time not spent doing something else. This paper explores what types of activities, and what specific activities within those types, compete with each other. For example, when you spend less time on social media, where do you spend more time? To answer these questions, we examine a 33 terabyte dataset of more than three trillion observations that includes information on every website that a panel of households visits over a period of four years. We sort these websites into categories of online activity. Among our key findings, social media and news are complements while social media and streaming are substitutes. This has implications for how we think about the markets for social media, news, and streaming, and the attention economy generally.

Keywords: antitrust, market definition, time use, competition

JEL Classification: K21, L12, L86, O33

‡ President and Senior Fellow, Technology Policy Institute, email: swallsten@techpolicyinstitute.org. The authors thank the Knight Foundation for financial support and Yoojin Lee and Nicolas Powell for research assistance. Any mistakes are our own.
* Senior Fellow, Technology Policy Institute, email: slam@techpolicyinstitute.org.
§ Senior Research Associate, Technology Policy Institute, email: nlovin@techpolicyinstitute.org.

I. INTRODUCTION

Our time is a scarce resource. We have only a limited amount, and today we spend a lot of it online for work, entertainment, shopping, social media, and more. Multitasking notwithstanding, when we are paying attention to one type of content, we are not paying attention to others.

The monetary price to consumers of many key online activities, like social media and search, is zero. Even with many popular services we may pay to subscribe to—video and music streaming, for example—we pay a fixed monthly fee and zero marginal dollars as we use it.

In a zero-price environment, which characterizes much online time, firms compete for people's attention. This type of competition has come to be known as the "attention economy." The attention economy poses problems for policy questions that involve market definition, including how to think about antitrust issues like mergers, where much traditional analysis focuses on prices and price competition. But when the price is zero, it becomes more difficult to evaluate competition.

To be sure, firms in the attention economy do compete in other traditional ways, even if not on price. They compete on dimensions such as quality, features, and bundles. But measuring competition in a market with zero (marginal) prices is arguably a key unique complication of the attention economy.

At a high level the concept is simple. If doing more of one thing causes you to do less of another thing, then those two activities are arguably substitutes and therefore compete with each other. Sometimes this makes intuitive sense. If viewing more TikTok videos results in less time on Facebook, it is not a stretch to conclude TikTok and Facebook compete. Switching from social media to streaming video might also sensibly lead one to conclude that they compete with each other. Online and offline activities can also compete with each other, but that is beyond the scope of this paper. [1]

In this paper we measure how Americans spend their time in the online attention economy and what activities are complements and substitutes. Using a massive dataset of home internet from Comscore's Total Home Panel we can determine what types of online activities households engage in at any moment in the day. We get a picture of the typical U.S. household's internet activity, but also which websites they tend to spend more and less time on in each category of online activity.

Our key conclusions are that even when instrumenting for endogeneity and estimating a simultaneous equation model, we find that time spent on social media and news are complements while time spent on social media and streaming are substitutes.

[1] Exploring online versus offline tradeoffs is difficult for reasons beyond data availability. What would it mean, for example, if we found in some hypothetical dataset that people switched from social media to an IRL ("in real life") activity such as hiking or cycling? It might be difficult to conclude that social media competes with hiking in a market sense, but in principle if that relationship were true then time spent on social media could be correlated with the IRL activity. For instance, social media or shopping time could relate to demand for, say, hiking boots. Whether online time and offline time compete or complement each other is a more challenging inquiry, since presumably time spent on shopping for hiking boots could vary per person and at some level would become competitive over complementary if the person spent more time picking out hiking boots than actually hiking. For previous work on this question, see generally Wallsten (2011). Wallsten, "What Are We Doing When We Are Not Online?," Economic Analysis of the Digital Economy (eds. Avi Goldfarb, Shane Greenstein, and Catherine Tucker), University of Chicago Press, https://www.nber.org/chapters/c13001.pdf.

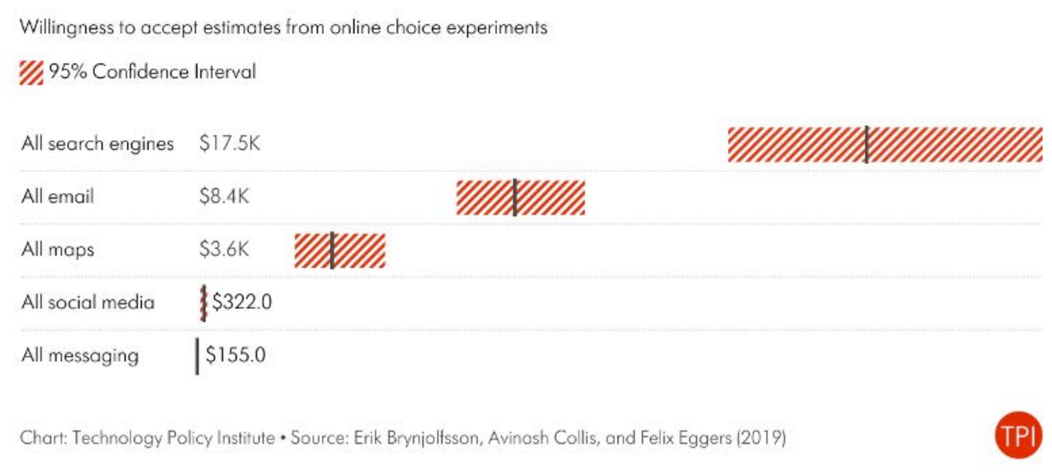
II. COMPETING FOR TIME

The first use of the phrase "attention economy" we can find is from a 1997 conference at Harvard's Kennedy School on the "Economics of Digital Information" by Michael H. Goldhaber.[2] He noted, "There is something [other than information] that moves through the Net, flowing in the opposite direction from information, namely attention. So seeking attention could be the very incentive we are looking for."[3] In other words, he correctly foresaw the enormous value of simply getting people to pay attention rather than to pay with money.

A large literature has explored how to measure the value of nonmarket goods.[4] A smaller body of literature has tried to estimate the monetary consumer value of "free" online services. Most research on the question has used the value of time as a measure of the value spent on various online activities. Goolsbee and Klenow (2006) estimated consumer value by using the prevailing wage rate and the time spent online.[5] This approach can provide a minimum value, as people would not spend that amount of time on something if it were worth less than the value of that time. However, it does not fully capture consumer welfare, as people may value the activity by more than just the time spent on it.

Research by Brynjolfsson, et al. (2019) builds on previous work by estimating how much people would have to be paid to voluntarily give up access to certain services. Figure 1 shows their results.

Figure 1: Per Consumer Value of Select Internet Services, 2017 (Brynholfsson, et al.)



We also know that the value of any activity to a person must be worth at least as much as the opportunity cost to them of doing the activity. That is, the value of any activity must be worth at least as much as the next-best thing they could be doing. Wallsten (2015) studied this aspect of online time using the American Time Use Survey to estimate how online time affected offline activities. [6]

[2] Michael H. Goldhaber, "The Attention Economy and the Net," First Monday, April 7, 1997, https://doi.org/10.5210/fm.v2i4.519.
[3] Goldhaber, id.
[4] See, e.g., Smith, V. Kerry. "Nonmarket Valuation of Environmental Resources: An Interpretive Appraisal." Land Economics 69, no. 1 (February 1993): 1. https://doi.org/10.2307/3146275.
[5] Goolsbee and Klenow (2006). Goolsbee and Klenow, "Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet."
[6] See, e.g., Wallsten (2015). Wallsten, "What Are We Not Doing When We're Online?"

Now, in this paper, we build on previous work by taking seriously the concept of "spending time" as a measure of value and using that to define markets. At a high level, products are competitors when they are substitutes. In our framing, online activities are substitutes when doing more of one directly leads to doing less of another, and vice versa.

A challenge is choosing the types of online activities likely to affect other activities. There is no hard and fast rule to do this. However, many sources tend to suggest that news, social media, and streaming all occupy a similar space online.

For example, social media is a source of key source of news for many. A Pew Research Center survey found that 71% of Americans obtain news content through social platforms, with Facebook being the most popular among them (Hutchinson, 2021).[7] Shearer (2018) finds that social media surpassed print newspapers as a primary source of news for younger people, with 36% of adults aged 18-29 reporting frequent news consumption through social media.[8]

That research suggests that time spent on social media and news can be both competitive and complementary. Social media sends people to read news, but the time spent on social media is less time spent on news if the user reads non-news content instead. Our empirical analysis makes an effort to tease apart these two phenomena.

III. DATA

We empirically investigate a detailed dataset of household internet use from Comscore. Specifically, we use data from Comscore's Total Home Panel that includes every website visited by every device in a representative sample of American households from every other month from September 2016 to November 2017 and August 2019 to June 2020.[9] Each observation in the data captures an instance a household device accessed a domain, along with descriptive information such as the URL accessed, services used, and device information. Each household is weighted so that the full monthly sample is representative of American households. Additionally, the data includes demographic information on each household, such as household size, number of children, income category, and zip code. The raw dataset includes about 33 terabytes of information and three trillion individual device-URL interactions. [10]

The nature of the raw data requires significant preparation to enable our crowding-out analysis.

[7] Hutchinson, Andrew. "New Research Shows that 71% of Americans Now Get News Content via Social Platforms." SocialMediaToday. January 12, 2021. https://www.socialmediatoday.com/news/new-research-shows-that-71-of-americans-now-get-news-content-via-social-pl/593255/

8 Shearer, Elisa. "Social Media Outpaces Print Newspapers in the U.S. as a News Source." Pew Research Center. December 10, 2018. https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/

[9] We use data from every other month as a way to maximize the time period covered with our fixed data budget. Every website visit for nearly 10,000 households is tracked in the months of Sept 2016, Nov 2016, Jan 2017, Mar 2017, May 2017, July 2017, Sept 2017, Nov 2017, Aug 2019, Oct 2019, Dec 2019, Feb 2020, Apr 2020, June 2020.

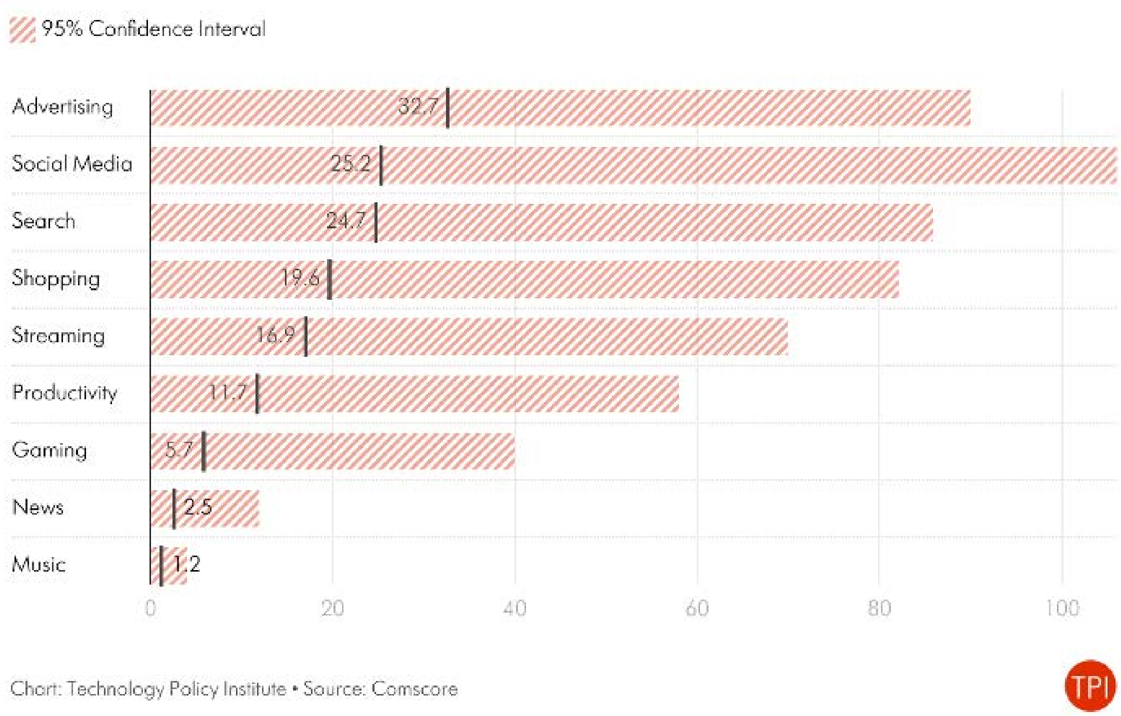
[10] We use Google BigQuery to manage 50 million rows of raw data and take a 5 percent random sample for analysis in Stata statistical software that can incorporate the household weights to the econometric analysis.

A. CATEGORIZING URLS

Our first step is to identify the categories of websites where household members spend time. This step is complicated by the hundreds of thousands of websites that household devices visit regularly. Thermostats, smart home speakers, connected lights, and even exercise machines check in frequently with cloud services. Some devices and domains are easy to identify and ignore. Most people don't spend a lot of time watching videos on their thermostats, for example. But others are more difficult. Websites almost constantly receive ads and send back other information in the background.

We determined a list of top domains by pulling the top 2,000 domains with the most visits across the panel of households on the first day of every month in the data. We sort domains into the following categories: advertising, gambling, gaming, music, news, pornography, productivity, search, shopping, social media, sports, streaming, torrenting, weather, and other. To do this, we used an algorithm followed by manual checking to categorize those domains, first using Cyren's URL Categorization Engine, then using research assistance.[11] If Cyren did not categorize a URL, we consulted secondary sites such as better.fyi and who.is. Figure 2 shows the time spent per hour.

Figure 2: Mean Minutes Per Hour Per Household Spent on Top Online Categories



On average, in a day in our dataset, a household interacted with 527 domains, or 202 domains per person given the weighted household panel composition of 2.6 people per household on average. This includes domains that do not contribute to active time spent online, such as domains connected to passive devices such as home appliances.

[11] https://www.cyren.com/products/web-security-engine

B. ESTIMATING TIME SPENT ON URLS

Obtaining the time spent on a site or online activity from the raw data required making some assumptions, since websites queried and data sent do not translate directly into units of time. In order to calculate how much time is spent on each domain, we assume that an average of two minutes is spent for each website access request. This assumption takes into account the various types of domains accessed by any given device, from streaming sites (which may send larger data packets that could be viewed for longer than two minutes) to social media sites or tracking domains (which may send smaller data packets that represent less than two minutes of viewing time).[12]

We round every observation down to the nearest two minutes and count each domain only once within the same two-minute period to avoid overcounting.[13] This time frame is not rolling, so a domain accessed at 12:01 and 12:02 will be counted twice, whereas a domain accessed at 12:00 and 12:01 will only be counted once. As a result, every count in the data is multiplied by two minutes in order to calculate the amount of time spent in each domain.

In a two-minute period, we see that multiple websites may be accessed at the same time, indicating multi-tasking. Thus, by our method, if we add up time spent over the day without accounting for multi-tasking, we would see online time far exceed 24 hours in a day. We spent considerable time with spot checks to see how the data stacked up in two-minute increments, and what type of web traffic and data packets were being sent to households.

Finally, because we are interested in peoples' behavior, we need to aggregate the data up to a level that reflects user decisions. To that end, we aggregate the data to the household-hour level, with separate variables for the time spent in minutes on each URL. This dataset has about 50 million household-hour observations.

The following section details what the data say about how we spend our time online.

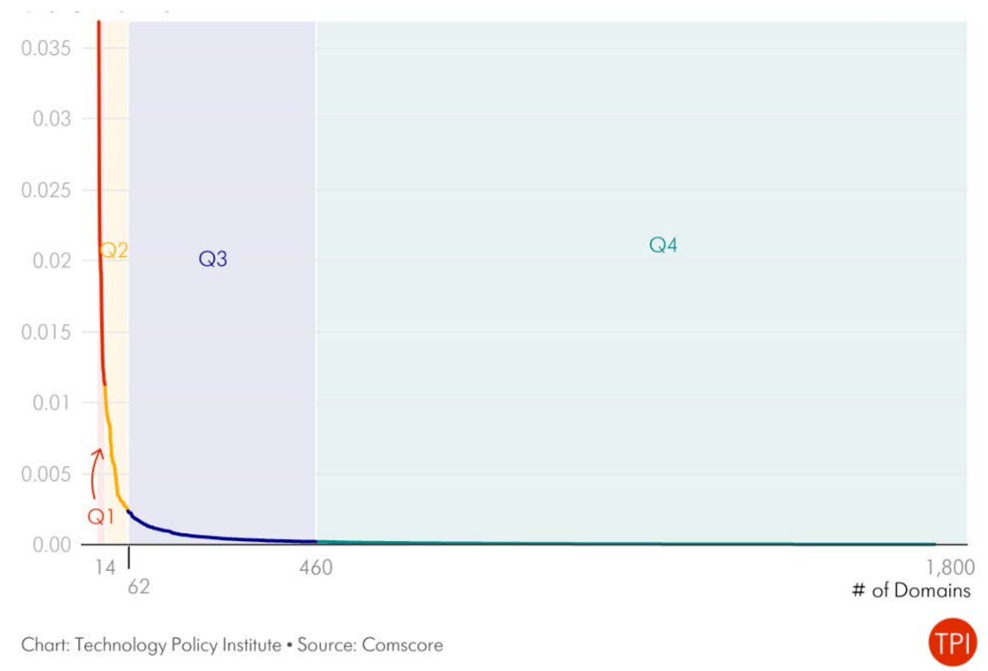
IV. HOW AND WHEN DO WE SPEND OUR TIME ONLINE?

Online time appears to be characterized by a "fat head" and "long tail" (Figure 3). From our list of top domains, 14 domains account for 25% of all time spent online. At the other end of the distribution, 1,340 domains also account for 25% of time spent online.

[12] As a robustness check, we applied different time units for each website access request, such as 5 minutes and 10 minutes. Since we were able to use big data tools, we counted smaller increments of 2 minutes for grouping the raw data by time. Time stamps in the Comscore dataset are demarcated at the millisecond level, but it would be cost-prohibitive to measure time at such small increments. While changing the length of time units has some effect on the estimated number of minutes on different categories, it does not change the qualitative results.

[13] The timestamp for each website access event is bucketed into 2 minute increments (i.e. floor(minute/2)).

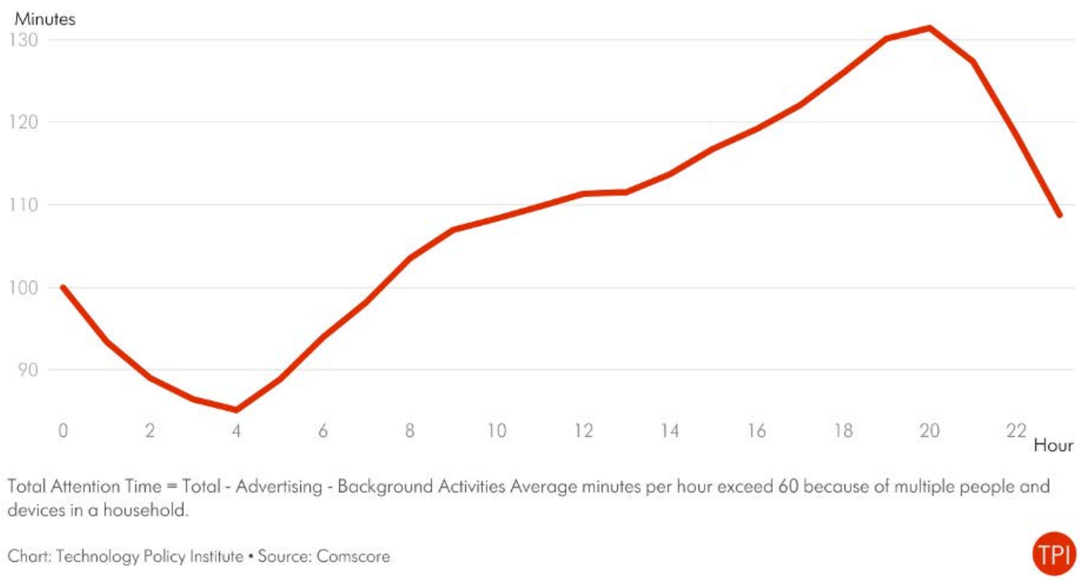
Figure 3: The Fat Head and Long Tail of Online Time, in Quartiles



A. ONLINE TIME THROUGHOUT THE DAY

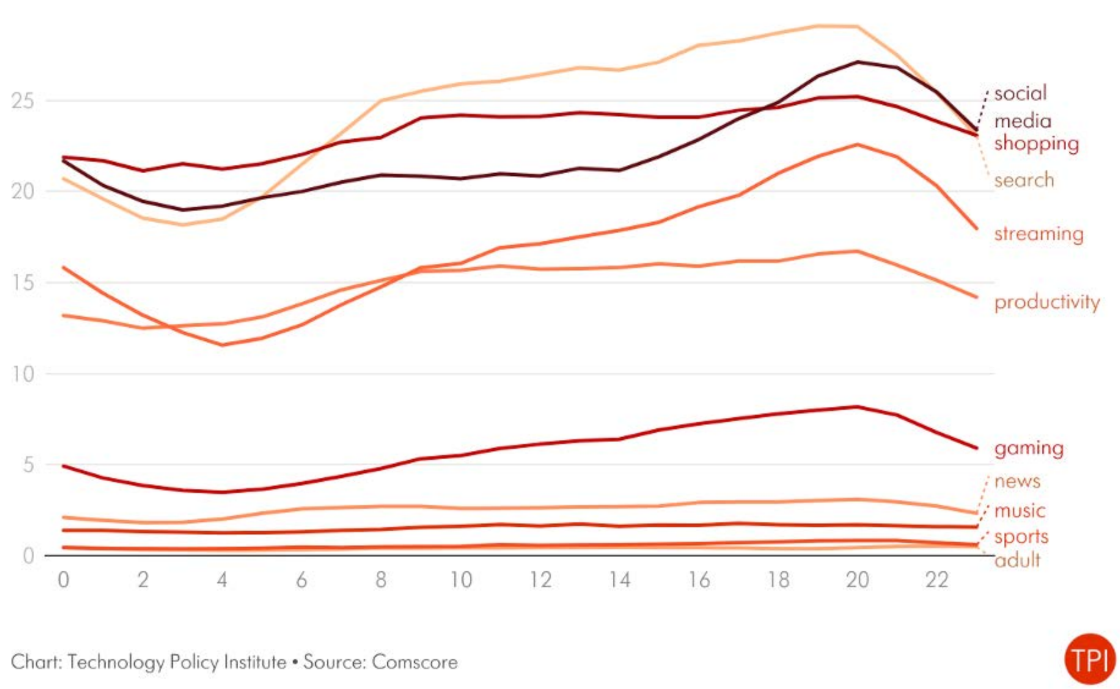
People do not spend their time online uniformly over time. Figure 4 shows that the peak of online time occurs at around 8:00 pm and the trough at around 4:00 am. As discussed above, the total number of minutes spent online in an hour exceeds 60 because it counts multi-tasking as separate activities and counts all members of the household.

Figure 4: Total Minutes of Online Time by Hour



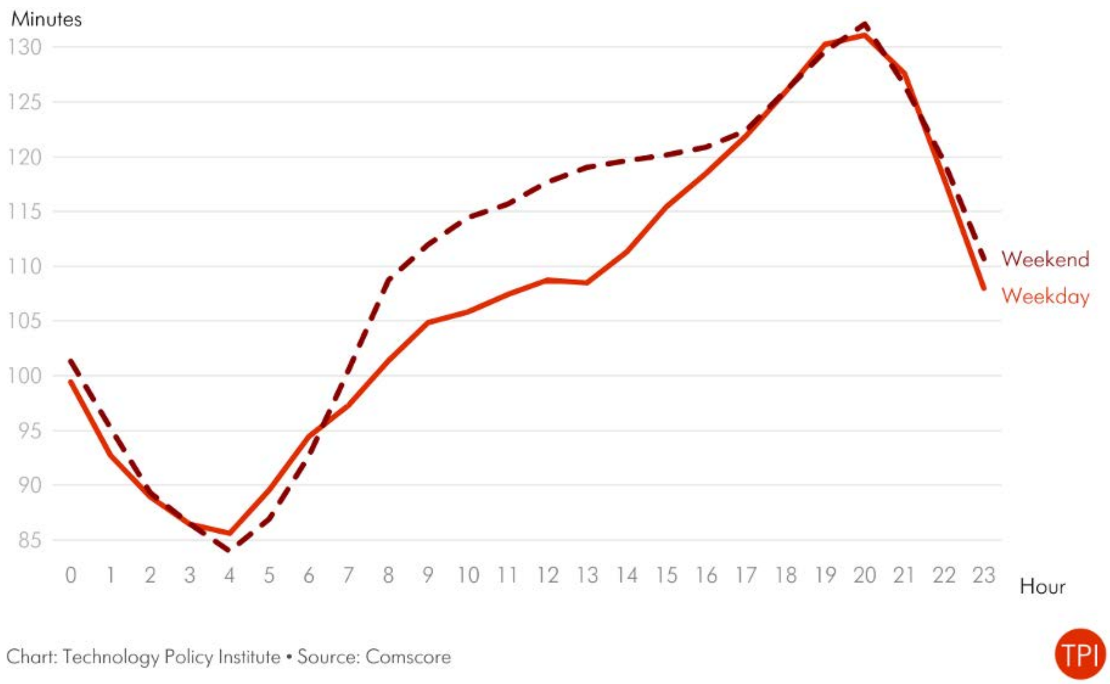
Viewing online time out by category shows some trend differences. Most peak and trough at 8 pm and 4 am, respectively. Shopping, however, is relatively stable across the day (Figure 5).

Figure 5: Minutes of Online Activities by Hour, All Days



Behavior on weekends differs from behavior during the week, with people spending more time online in the middle of the day on the weekend than they do during the week (Figure 6).

Figure 6: Weekday and Weekend Minutes Online by Hour



B. ONLINE TIME ACROSS YEARS

The amount of time people spend online over time changed during the time period we studied, generally increasing from 2016 through 2020. One striking, if not surprising, trend apparent in the data is the spike in online time when the Covid-19 pandemic hit. In the figures below, dates that include the pandemic are highlighted in gray.

Recall that to maximize the time period we could cover with a limited budget we obtained data only every other month. We combined newer 2019-2020 data with data we had obtained for a different project that covered months from 2016-2018.

Figure 7 shows average household daily time spent online for the two time frames. Total time increases over time with a notable increase observed during the pandemic months.

Figure 7: Average Household Daily Time Spent Online in Hours (2017-2018, 2019-2020)

Total exceeds 24 hours per day because households have multiple people.

Chart: Technology Policy Institute • Source: Comscore

C. ONLINE TIME ON SPECIFIC SITES WITHIN CATEGORIES

Certain activities appear to have become more and less popular over time. Time spent shopping steadily increased over this time period, with a large jump when the pandemic began (Figure 8). Search and productivity also increased steadily, particularly during the pandemic. Social media decreased over time.

Figures 8 through 17 show how people spend their time online by different dimensions and categories. We show time spent by online category, time spent on top social media platforms, time spent on top streaming platforms, time spent on major news sites, time spent on news by the hour of day, time spent on streaming by hour of day, time spent on productivity sites by hour of day, time spent on shopping sites by hour of day, time spent on gaming by hour of day, and time spent on gaming sites by teenage boys and girls in a household.

Figure 8: Minutes Per Day on Certain Online Activities (2017-2018, 2019-2020)

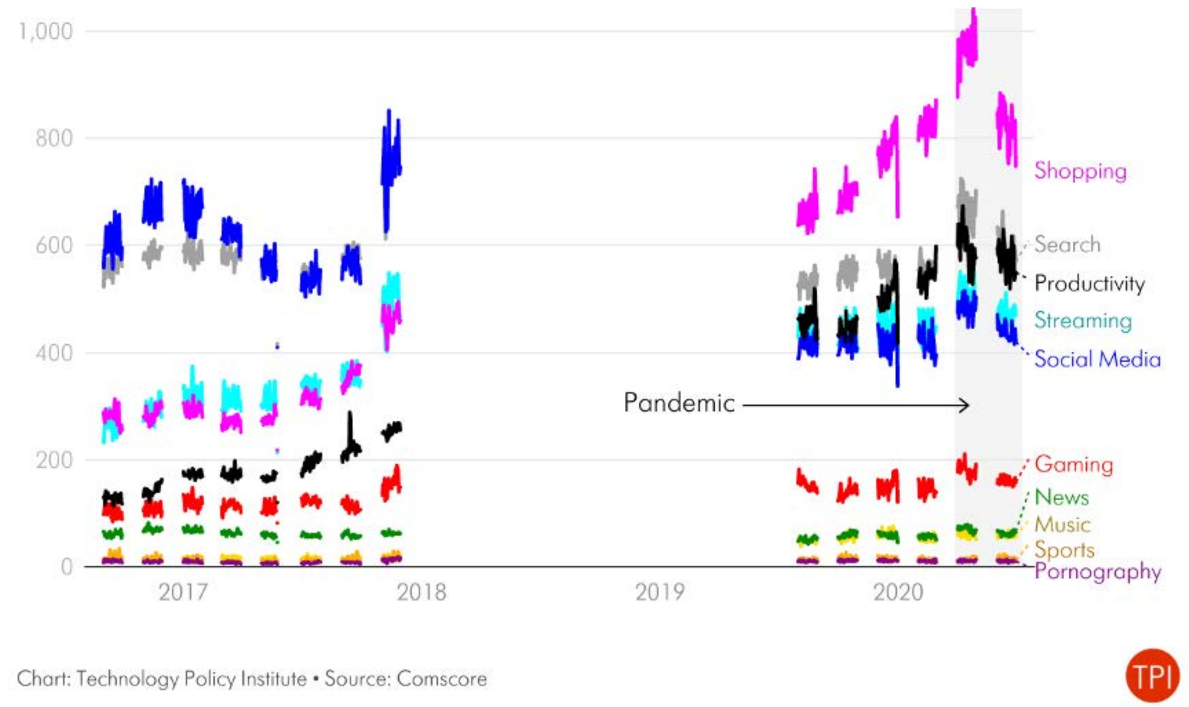


Figure 9: Minutes Per Day on Top Social Media Platforms (2017-2018, 2019-2020)

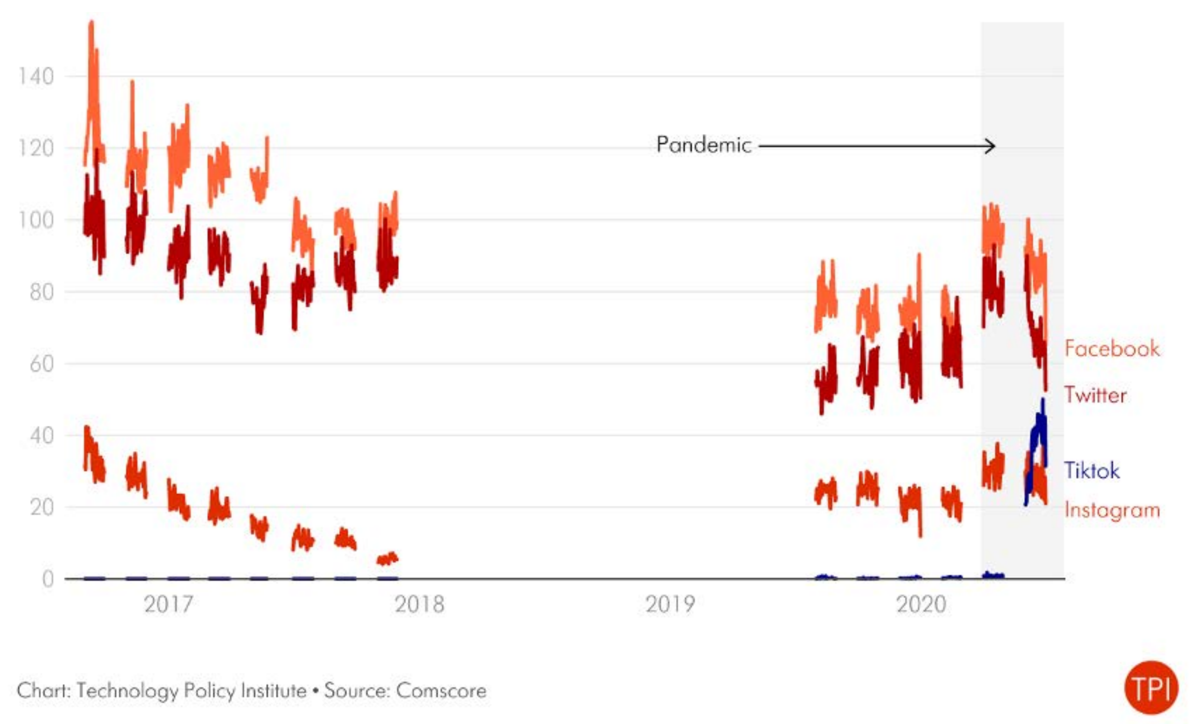


Figure 10: Minutes Per Day Streaming Over Time (2017-2018, 2019-2020)

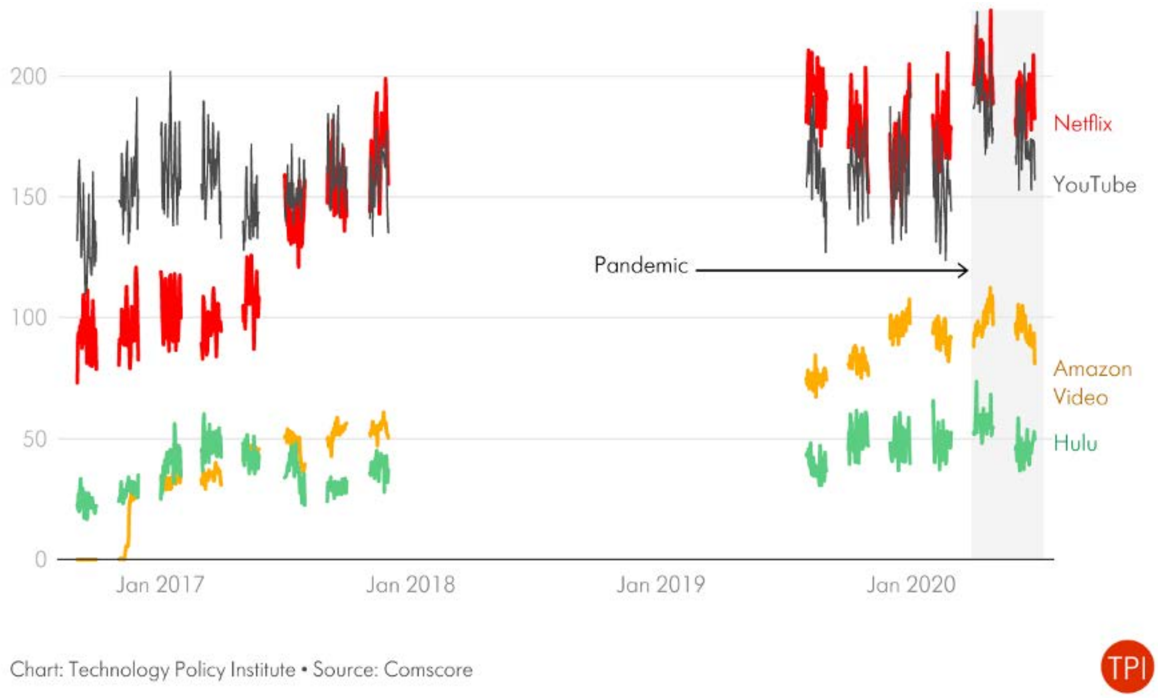
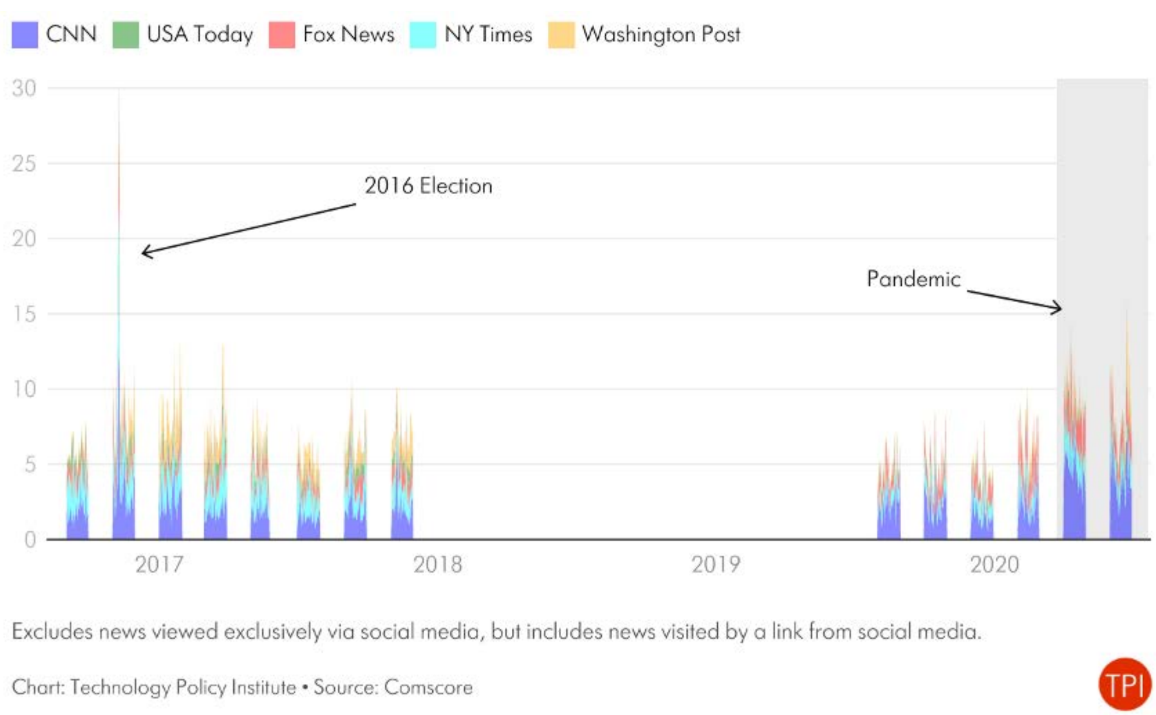


Figure 11: Average Minutes Per Day on Major News Sites (2017-2018, 2019-2020)



D. CHANGES IN ONLINE BEHAVIOR IN PRE-PANDEMIC AND DURING-PANDEMIC MONTHS

Figures 12 through 16 show changes in online behavior during the Covid-19 pandemic period with shutdowns and nationwide shifts to virtual school and work. Americans use of time during the day shifted across hours to more online time. We observe the change in news, streaming, productivity, shopping, and gaming categories. More notable shifts upward in online time during daytime hours are seen in news, productivity, and shopping, compared to pre-pandemic days. Streaming and gaming has an increase but not as notable.

Figure 12: Average Online News Time by Hour in Minutes

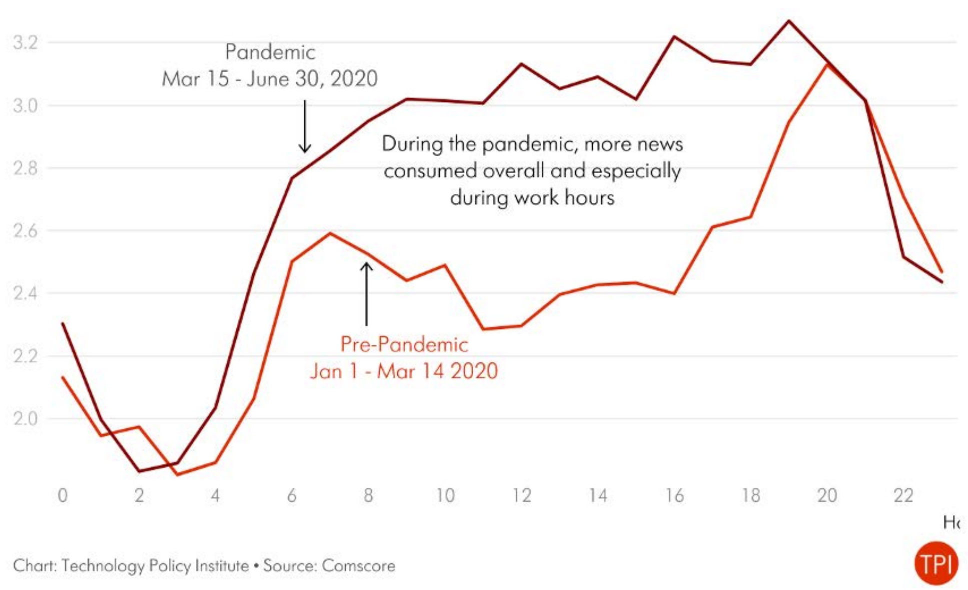


Figure 13: Average Streaming Time by Hour in Minutes

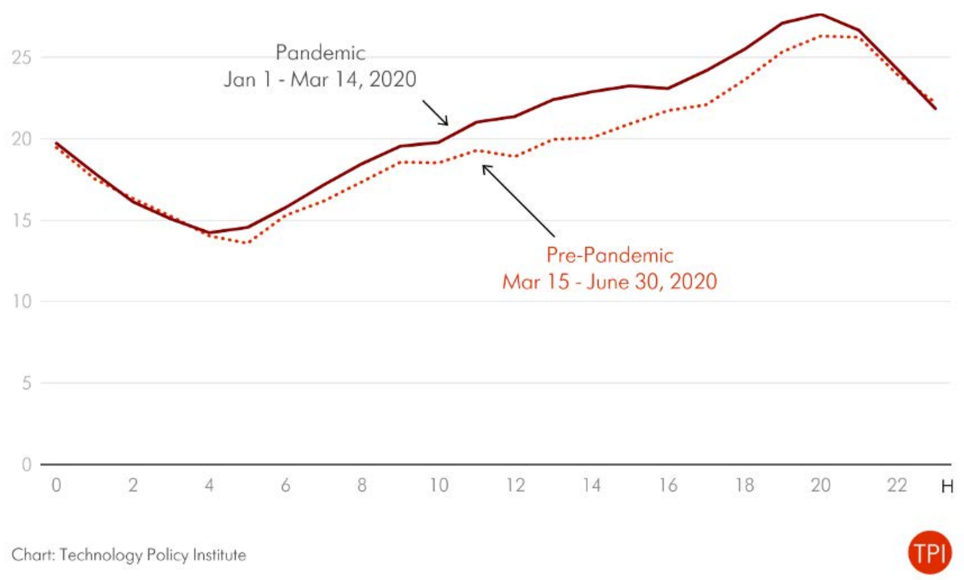


Figure 14: Average Time on Productivity Sites by Hour in Minutes

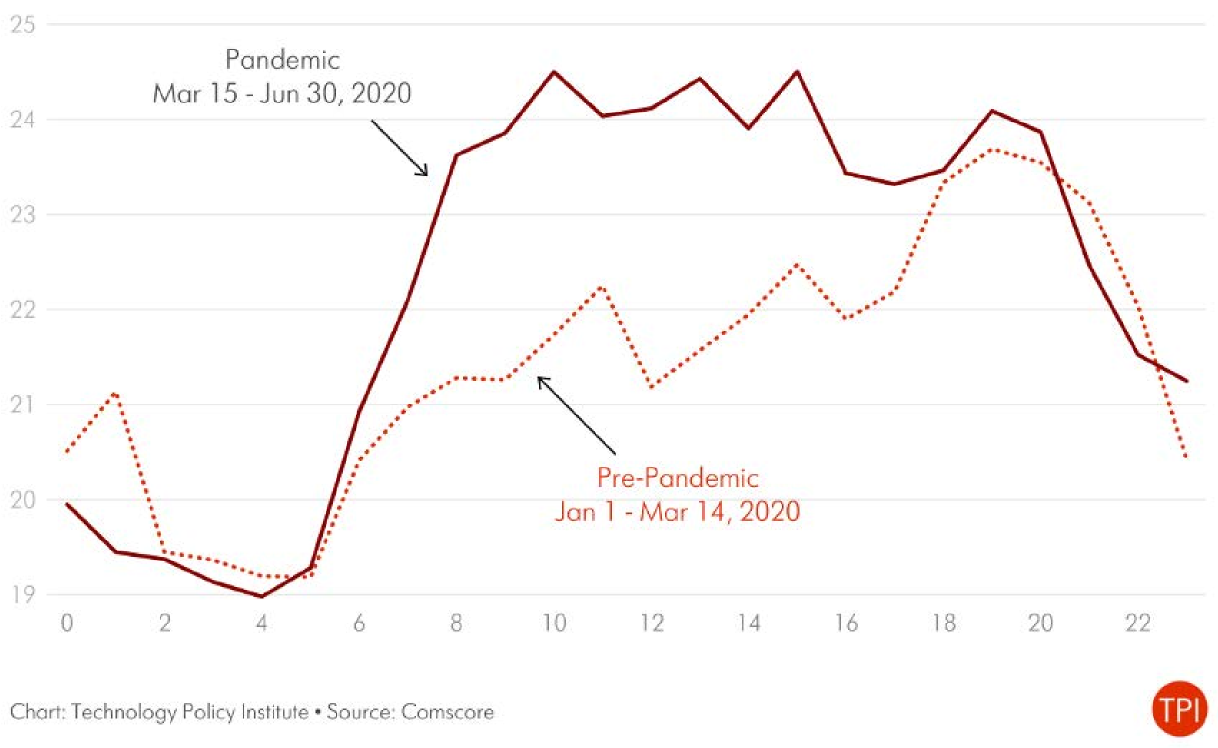


Figure 15: Average Online Shopping Time by Hour in Minutes

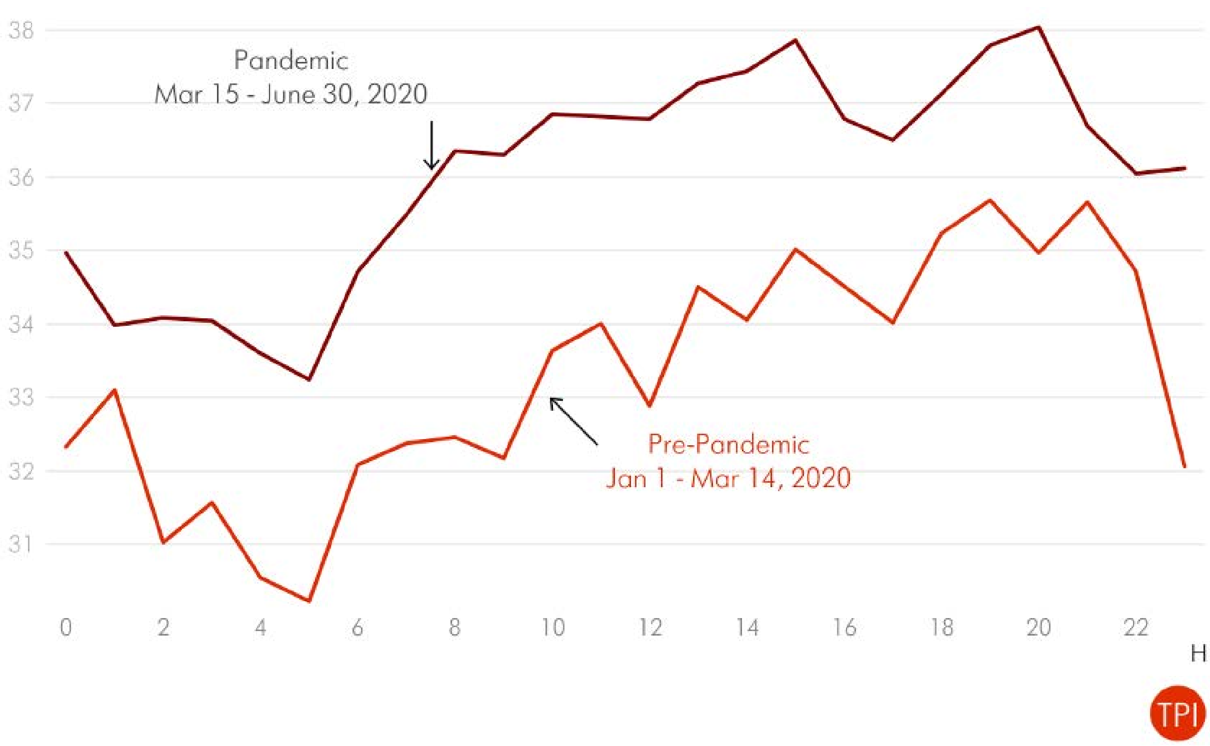
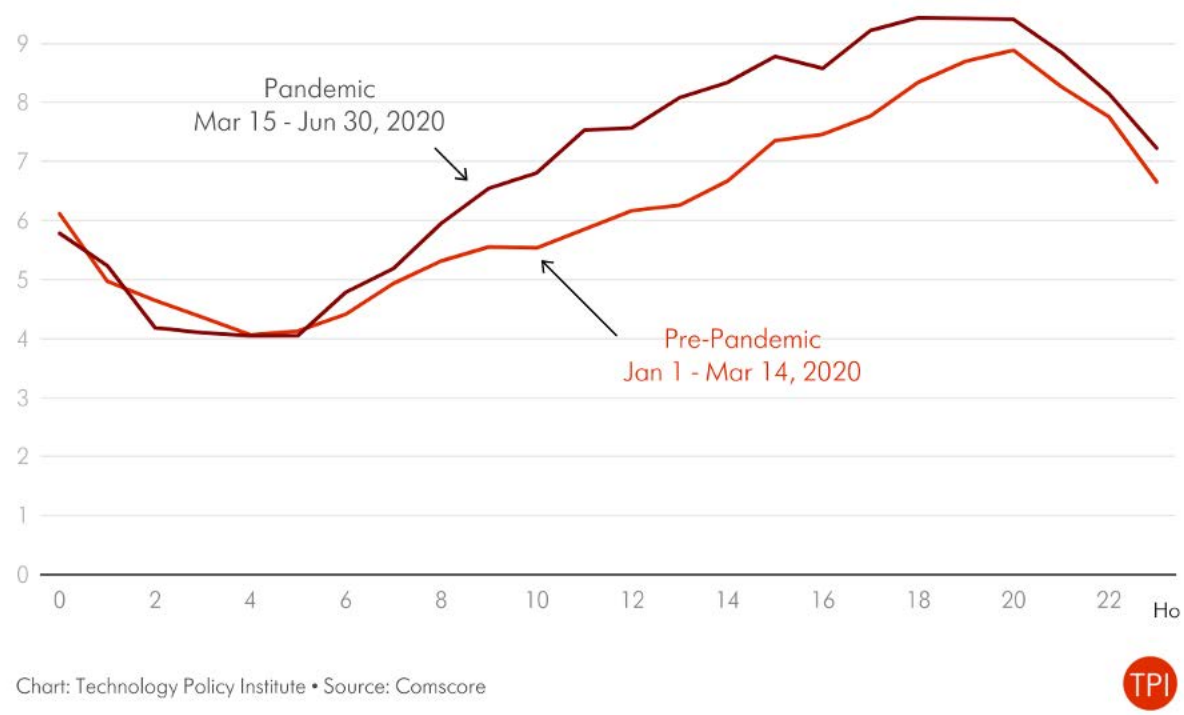


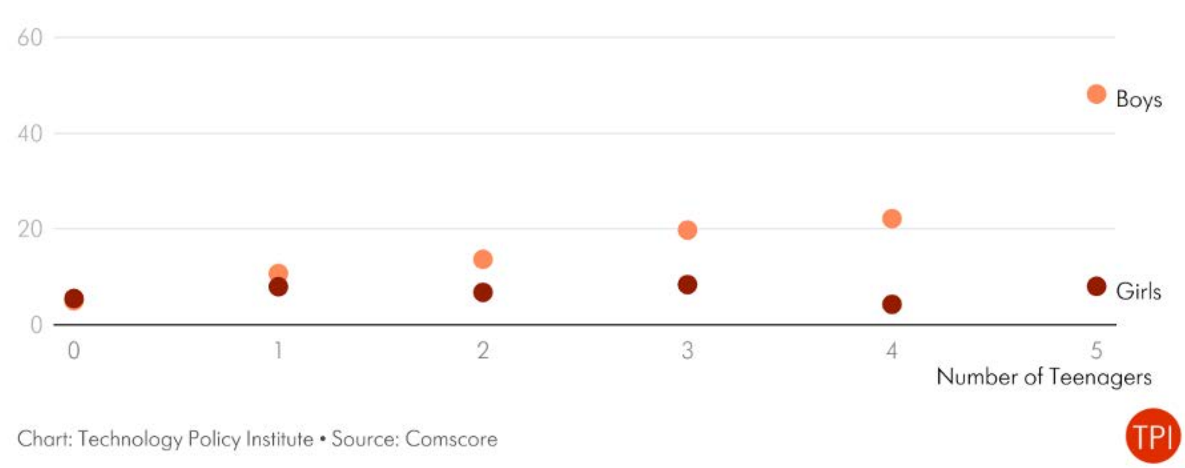
Figure 16: Average Gaming Time by Hour in Minutes



E. GAMING BY GENDER

Households with teenage boys spend more time on gaming websites than households with teenage girls as seen in Figure 17. As the number of teenagers in a household increases from 0 to 5, there is a more noticeable trend that boys in the household spend more minutes on gaming sites than households with multiple girls. Our study does not look at gender effects in particular, but we include demographics such as teenage girls and boys in the econometric analysis explained in the next sections.

Figure 17: Minutes Gaming by Number of Teenagers by Gender in Household



V. SUBSTITUTES AND COMPLEMENTS

When defining markets in attention, the key insight is to think in terms of crowding out—how does time spent on one site affect time spent on another?[14] The dependent variable is the time spent on one site (or on one category of online activity) and the independent variable is time spent on a separate site (or a different category of online activity). The key empirical challenge is finding an instrumental variable to correct endogeneity problems that causes the selection of one activity over another. This endogeneity problem is explained in closer detail.

A. MARKET DEFINITION AND CROWDING OUT

In a standard antitrust analysis, a first step is typically defining the market, which requires evaluating how demand for a given product and demand for potential complementary and substitute products change as the price of the product changes (i.e., price and cross-price elasticities). In the market analysis, the dependent variables are quantity of the products sold, which is typically measured in revenue, units shipped, or some other metric that reflects quantity. The independent variable of interest is the price set for the product in question.

In the attention economy, the dependent variable is also a quantity of products sold, but since most online content has a zero price and is not sold, we measure instead the quantity of time spent by the user on a given site. The market is best defined by how users allocate their time between alternative sites. For this market definition exercise, the empirical analysis focuses on crowding out (or in) effects. That is, does spending time on TikTok cause less time spent on Facebook? At a high level, the dependent variable would be time spent on Facebook and the independent variable time spent on TikTok. Estimating that effect empirically would require finding an instrument for TikTok.

A simple crowding out test like that, however, is problematic because time spent on the two sites is endogenous—we don't know the direction (or necessarily even source) of causality. In some cases, we can determine causality from timing or other factors. For example, if TikTok enters in 2019 and we observe a decline in time spent on Facebook that corresponds with an increase in time spent on TikTok both in timing and magnitude, it is likely that TikTok is crowding out Facebook.

Empirical analysis is never that clear, however. The key will be finding the right instrumental variable that econometrically identifies the variable we want to treat as independent. A valid instrument is a variable that is correlated with someone choosing to use TikTok but uncorrelated with that person choosing not to use Facebook (except via the TikTok tradeoff). [15]

---

[14] See generally Wallsten (2015). Wallsten, "What Are We Doing When We Are Not Online?," Economic Analysis of the Digital Economy (eds. Avi Goldfarb, Shane Greenstein, and Catherine Tucker), University of Chicago Press, https://www.nber.org/chapters/c13001.pdf.

[15] In a previous TPI paper on video streaming piracy, which had a similar crowding-out approach, we identified a household's decision to watch pirated videos instead of non-pirated videos. The best instrument for identification, it turned out, was whether a household had a PC running Windows (as opposed to a Mac or only mobile devices or connected television). Those households were much more likely to pirate at least some of their video than households without PCs.
The piracy example highlights how an instrumental variable is supposed to work in theory as well as the complications of using it in practice. This instrument worked econometrically and passed various statistical validity tests. However, not everyone who owns a PC pirates video and not everyone who owns a Mac does not. Across our trillions of observations in the dataset this instrument nevertheless helps to identify the crowding out effect. We will have to find similar instruments as we empirically study the market definitions. https://techpolicyinstitute.org/wp-content/uploads/2020/01/Oh_Wallsten_Lovin_Streaming-Video-Piracy.pdf

B. SIMPLE CORRELATIONS

The figures above show what people do online and when, but do not directly examine which online activities are complementary and which are substitutes.

A first step is to look at correlations between each variable pair. Figure 18 shows the correlation matrix, color coded to aid interpretation.[16]

Figure 18: Pairwise Correlations Between Online Categories (abridged)

	Social Media	News	Gaming	Streaming	Shopping	Music	Productivity
Social Media	1.000						
News	0.211	1.000					
Gaming	0.174	0.065	1.000				
Streaming	0.245	0.095	0.188	1.000			
Shopping	0.156	0.101	0.079	0.222	1.000		
Music	0.061	0.032	0.070	0.074	0.070	1.000	
Productivity	0.182	0.133	0.104	0.189	0.644	0.095	1.000
Search	0.552	0.237	0.188	0.301	0.206	0.091	0.247

Nearly every activity is positively correlated with every other activity, although in most cases the correlation coefficient is fairly small in magnitude. The generally positive correlation is not surprising: pairwise correlations do not control for any other factors, and it is likely that if people in a household generally do a lot of one thing online they probably do a lot of other things online.

Still, some of the pairwise correlations suggest interesting relationships. The strongest correlation is between productivity and shopping. It says nothing about causality, but it suggests that people tend to shop while they work. We do not know whether they shop because they are spending time working, or whether they are using productivity tools because they are also shopping. We also do not know whether they would be working more if they were not also shopping, or if they would be shopping more if they were not also working online. We just know that these two online activities move directionally together. The next strongest correlation is between social media and search. Social media generally is strongly correlated with other online activities, perhaps because people who spend more time on social media spend more time online generally.

C. ECONOMETRIC METHODOLOGY

In general, as the correlation matrix shows, time spent on one activity is positively correlated with all other activities. This is not surprising—households that spend more time online generally spend more time on everything online. We therefore control for other factors likely to influence time spent online, including household size, number of teenage boys, number of teenage girls, income, number of connected devices, time of day, day of week, month, and year. To this we add broadband connectivity (from the FCC's broadband map): the share of households in the zip code with access to broadband service offering at least 100 Mbps download and 20 Mbps upload. Households that can subscribe to faster tiers of broadband service

[16] The matrix is abbreviated to show the most relevant information, the full correlation matrix of all categories is available on our website.

presumably have a greater capacity to multitask and spend more time online. Note that because this variable measures connectivity at the zip code level it does not have the problem of the household selecting faster speeds because they want to consumer more online content. [17]

These controls are not sufficient, however. The "independent" variables we care about are also endogenous. For example, we want to know how time spent on social media affects time spent reading online news. News cannot properly be considered exogenous to social media because social media is often the gateway to news stories and trending topics. We need a strategy to identify the effects of time spent in certain areas.

We instrument for streaming, social media, and news consumption. We identify streaming by the number of streaming devices in a household, news by the number of newspapers per capita in the household's state,[18] and social media by the number of phones in a household.

Endogenous Variable	Instrument
Streaming Minutes	Number of streaming devices in household
Social Media Minutes	Number of phones in household
News Minutes	Number of newspapers per hundred thousand people in the state

We explore the relationship between social media and news, social media and streaming, and streaming and news. For each pair, we estimate a simultaneous equation model as in equations (1) and (2) below using three-stage least squares. The model identifies both endogenous variables.

$$\begin{aligned}\text{social\_media}_{ht} =&\beta_0 + \beta_1\widehat{\text{news}}_{ht} + \beta_2\text{num\_phones}_{ht}\\ &+ \beta_3\text{num\_devices\_less\_phones}_{ht} + \beta_4\text{hh\_size}_{ht}\\ &+ \beta_5\text{num\_kids\_not\_teens}_{ht} + \beta_6\text{num\_teenage\_boys}_{ht}\\ &+ \beta_7\text{num\_teenage\_girls}_{ht} + \beta_8\text{share hhlds in zip with access to 100/20 broadband}_{ht}\\ &+ \beta_9\text{income}_{ht} + \beta_{10}\text{hour o'clock}_{ht}\\ &+ \beta_{11}\text{day of week dummies}_{ht} + \beta_{12}\text{month dummies}_{ht}\\ &+ \beta_{13}\text{year dummies}_{ht} + \epsilon_{1ht}\end{aligned} \tag{1}$$

$$\begin{aligned}\text{news}_{ht} =&\alpha_0 + \alpha_1\widehat{\text{social\_media}}_{ht} + \alpha_2\text{num\_devices\_less\_phones}_{ht}\\ &+ \alpha_3\text{hh\_size}_{ht} + \alpha_4\text{num\_kids\_not\_teens}_{ht}\\ &+ \alpha_5\text{num\_teenage\_boys}_{ht} + \alpha_6\text{num\_teenage\_girls}_{ht}\\ &+ \alpha_7\text{share hhlds in zip with access to 100/20 broadband}_{ht}\\ &+ \alpha_8\text{newspapers per hundred thousand ppl}_{ht} + \alpha_9\text{income}_{ht}\\ &+ \alpha_{10}\text{hour o'clock}_{ht} + \alpha_{11}\text{day of week dummies}_{ht}\\ &+ \alpha_{12}\text{month dummies}_{ht} + \alpha_{13}\text{year dummies}_{ht} + \epsilon_{2ht}\end{aligned} \tag{2}$$

Recall that our household-hour level dataset contains more than 50 million observations. It is both impractical and unnecessary to estimate these equations across the entire dataset. We therefore take a five percent random sample and estimate the equations with a smaller dataset of 2.6 million observations.

[17] Except, of course, to the extent that a household's desire to live in a zip code with higher connectivity overwhelms other factors affecting where to live.
[18] Newspaper circulation data comes from the University of North Carolina at Chapel Hill. State population data comes from ACS.

The following section shows the results of estimating these equations.

VI. RESULTS

The following tables show the results for each of the three pairs of equations.

Table 1 shows the results of estimating the system exploring the relationship between news and social media. All coefficients are statistically significant, which is to be expected with a dataset of about 2.6 million observations. The first two rows show the variables of interest. The table shows a positive relationship between news and social media in both directions: additional time on social media translates into more time on news and vice versa.

Table 1: Social Media and News: Simultaneous Equations Regression Results

		Dependent Variable	
Variable	Mean	Social Media Minutes	News Minutes
News Minutes	2.6	18.32	
Social Media Minutes	22.4		0.12
Number of phones	7.7	-0.03	
Newspapers per hundred thousand people in state	2.2		-0.04
Number of devices other than phones	-0.03	0.001	-0.04
Household size	2.8	0.53 (p=0.171)	-0.4
Number of kids < 13	0.23	3.43	0.09
Number of teenage boys	0.15	7.68	-0.06 (p=0.036)
Number of teenage girls	0.14	8.03	-0.13
Share of households in zip code with access to 100/20 broadband	-0.07	0.01	
Number of observations 2,681,295			

All coefficients statistically significant at $p<0.01$ unless otherwise noted.
Hour, day of week, month, income group, and year fixed effects included.
Estimated using 3SLS.

The results appear to be meaningful in terms of magnitude. Which has a "bigger" effect depends on whether we consider percent changes or minute changes. Each additional minute spent with online news generates another 18 minutes on social media. Each additional minute on social media generates 0.12 additional minutes on news.

Table 2 shows the results of estimating the equations measuring social media and streaming. In this case the results suggest that social media and streaming are substitutes.

Table 2: Social Media and Streaming: Simultaneous Equations Regression Results

		Dependent Variable	
	Mean	Social Media Minutes	Streaming Minutes
Streaming Minutes	16.5	-0.15	
Social Media Minutes	22.4		-0.28
Number of phones	7.7	0.02	
Number of streaming devices	2.1		0.21
Number of devices other than phones or streaming devices	19.2	0.01	0
Household size	2.8	5.85	3.43
Number of kids < 13	0.23	-3.93	0.20
Number of teenage boys	0.15	-4.83	1.44
Number of teenage girls	0.14	-4.06	1.72
Constant	22.16	9.06	
Number of observations 2,681,295			

Hour, day of week, month, year, and income group fixed effects included.
Estimated using 3SLS.

Each additional minute spent streaming leads to 0.15 minutes less time on social media. Similarly, each additional minute spent on social media leads to 0.28 minutes less time on streaming.

Table 3 shows the estimated relationship between streaming and news. Unlike the two discussed above, this relationship is more complicated. The results suggest that streaming crowds out news, but that news crowds in streaming. That is, watching more streaming reduces time spent reading news, but time spent reading news increases time spent streaming.

Table 3: News and Streaming: Simultaneous Equations Regression Results

		Dependent Variable	
	Mean	News Minutes	Streaming Minutes
Streaming Minutes	16.5	-0.02	.
News Minutes	2.6	.	7.02
Newspapers per hundred thousand people in state	2.2	0.04	
Number of Streaming Devices	2.13		0.25
Number of devices other than for streaming	19.2	0	-0.02
Household size	2.8	0.32	-0.04
Number of kids < 13	0.23	-0.38	4.23
Number of teenage boys	0.15	-0.64	7.85
Number of teenage girls	0.14	-0.62	7.75
Share households in zip with access to 100/20 broadband	54.2	0	-0.02
Constant		2.16	-12.32
Number of observations 2,681,295			

All coefficients statistically significant at p<0.01.
Hour, day of week, month, income group, and year fixed effects included.
Estimated using 3SLS.

A. CAVEATS

The results we present are robust to empirical specification. However, as we discussed above, a dataset this large and disaggregated required making many assumptions in order to make usable. Some of those assumptions might affect the results.

Probably the biggest source of potential error is the way we assign time to site visits. As discussed above, we count each site visit as two minutes of time. This approach likely works well for streaming, where the devices contact the site regularly as the person watches, but it might work less well on a news site, where someone opens a news story and reads it for a while.

Additionally, the categorization is not perfect. The obvious potential error here is miscategorizing a site, most likely because some sites do not fit neatly into a single category. Social media complicates the categorization further. If someone clicks from a social media site to a news story, for example, we capture time on the social media site and on the news site, but if the news story is embedded within a social media post, there is no referral traffic to an external domain and we would categorize the time as social media and not news.

We can ask dozens more questions of this dataset, such as device-level behavior, multi-tasking behavior, and time-of-day behavior as well. For the purposes of defining attention markets and crowding out effects, we developed a methodology for categorizing online sites and measuring time spent on the sites. Using these data cleaning steps and combining broadband data along with demographic data, we applied econometric methods to ask questions about whether online activities are complements or substitutes.

VII. CONCLUSION

Time spent online has become a significant part of our work, educational, and leisure activities. Yet, little research has examined how to think about the attention economy as markets in the sense of what activities are substitutes and complements. Research along these lines is crucial to help us understand competition in the attention economy, particularly with many activities requiring zero monetary payments by users.

In this paper we use a 33 terabyte dataset of every device-url interaction in a panel of households covering 2016 to 2020 from Comscore's Total Home Panel to investigate how people spend time online and how social media, news, and streaming interact with each other. Our primary results from estimating simultaneous equations and instrumenting for endogeneity are that news and social media are complements while social media and streaming are substitutes.

These results have implications for how we think about competition in social media, news, and streaming and how we might define markets in those activities.

TECHNOLOGY
POLICY
INSTITUTE

2001 L Street NW, Suite 500, Washington, D.C., 20036,
Website: https://techpolicyinstitute.org, Twitter: @techpolicyinst