Does Autoplay Drive Excessive Screen Time? Evidence from an Online Experiment

Reha Tuncer*

August 18, 2025

Abstract

Interface design features can 'nudge' consumers to take certain actions and are often accused of promoting addictive online behavior. A prevalent design feature across popular social media and streaming platforms is the autoplay default. In this study, I present an incentivized online experiment investigating whether the autoplay feature can cause an increase in undesired video consumption, and elicit the willingness to pay for commitment against autoplay. In a two-day study, I recruited a total of 236 participants to allocate 20 minutes between two tasks: Transcribing meaningless characters and watching funny animal videos. Time allocation decisions were planned a day before and realized on the next day. I randomly assigned participants to either autoplay or click-to-play media controls while keeping the video content constant. I find that the autoplay feature, in isolation, does not override participants' planned time allocation

*I am grateful to Ernesto Reuben, Kerstin Bongard-Blanchy, and Vincent Koenig for their guidance in the conception of this experiment, to Sophie Doublet for her help in interface design, and to Suvadeep Mukherjee for his guidance in building web experiments. I also thank participants at the Behavioral Economics Summer School at the University of Copenhagen, the doctoral workshop at UC Louvain, the Experimetrics Workshop Soleto at Sapienza University Rome, as well as Samuel Greiff, Manuel Munoz, Hande Erkut and numerous colleagues for helpful comments. This study is supported by the Luxembourg National Research Fund (FNR) PRIDE 19/14302992. All errors remain my own.

for media consumption. In addition, participants exhibit a positive willingness-to-pay for autoplay (6.72 pence/hour), perceiving it as a convenience feature rather than a self-control problem. Experimenter demand effects and lack of content appeal result in participants allocating more time to the transcription task than planned, confounding the effect of the autoplay treatment. These results suggest that design features promoting potentially addictive behavior like autoplay are better studied in field settings where content consumption occurs naturally alongside algorithmic personalization.

JEL Classification: C91, D12, L82, D90

Keywords: experimental economics, digital addiction, interface design, self-control, media consumption

1 Introduction

Autoplay is a design feature that automatically provides users with new video content without any action. It has become a familiar sight in the user interfaces of major social media and streaming platforms over the last decade,[1] and has been described as a default nudge that reduces the autonomy of users (Lukoff et al., 2021; Schaffner et al., 2023). Autoplay's widespread adoption and concerns about its restriction of users' freedom of choice have also attracted regulatory scrutiny, with policymakers from both sides of the Atlantic proposing to ban platforms from serving autoplay content due to its contribution to addictive online behavior.[2]

While autoplay is pervasive and tends to increase viewing times (Chen et al., 2025; Hiniker et al., 2018; Schaffner et al., 2025), its impact on user welfare and the mechanisms driving this effect are not well understood. Does the increase in consumption go against users' wishes, or does it merely remove frictions to help them achieve their

[1]As of 2025 these platforms include Facebook, Instagram, Netflix, YouTube, TikTok, and Twitter (currently X). Twitter first introduced autoplay in 2015, and described it as a means to reduce "extra effort" in the number of clicks/taps required to consume content.

[2]See US SMART Act of 2019, S. 2314, 116th Congress, and the upcoming EU Digital Fairness Act currently in discussion.

desired consumption levels? Can we attribute the increase in consumption to autoplay in isolation, or does it also depend on the supply of content? To isolate the causal effect of autoplay on content consumption and its welfare consequences, I designed an experiment holding constant the video content displayed to users while eliciting their preferences related to content consumption before the actual consumption occurs.

I dedicated the first block of the experiment to measure the differences between planned and realized allocations. I recruited 184 participants who allocated 1200 seconds (20 minutes) between two tasks. The tasks involved seconds of transcription of random character sequences at 0.15 pence/second, and seconds of watching funny animal videos at 0.1 pence/second. On the first day, participants got familiar with tasks and made planned allocations using the strategy method. On the following day, participants realized their time allocation decision without constraints.

To isolate the causal effect of autoplay on the realized time allocation decisions, I randomly assigned participants to either the *Autoplay* or *Control* condition. In the *Autoplay* condition watching task videos played continuously. In the *Control* condition participants explicitly clicked to play each video. I held constant the content by presenting 80 manually curated "funny animal" videos from YouTube and TikTok in identical sequence. This eliminated algorithmic content curation and personalization as potential confounds.

In a second block of the experiment, I modified the design to elicit demand for a commitment device. I recruited 52 additional participants who made binary choices over their preferred autoplay setting (on or off) for nine bonus payments ranging from 5 to 50 pence. I implemented one of these decisions at random, determining both the autoplay setting for the next day and the associated bonus payment. This allowed me to measure participants' Willingness To Pay (WTP) for autoplay in advance, providing insight into whether participants viewed the feature as beneficial (positive WTP) or as a self-control problem requiring a commitment device (negative WTP).

I report five key findings. The first set of results focuses on my preregistered[3] hy-

[3] Access the preregistered hypotheses.

pothesis that autoplay would cause an increase in video content consumption. I find no evidence supporting this hypothesis. Participants in the *Autoplay* condition watched statistically indistinguishable numbers of videos compared to the *Control* condition. *Autoplay* condition also had no significant effect on the proportion of time spent on either task. These results also hold for other engagement metrics (i.e., average transcribing/watching session length, or the number of transcribing/watching sessions).

Second, participants viewed autoplay as beneficial rather than problematic. Contrary to my hypothesis that participants would demand commitment devices against autoplay, I find a positive WTP for the feature. Using data from the second block, I estimate an average WTP of 6.72 pence/hour. Autoplay is perceived as a convenience feature rather than as a threat to self-control, aligning with earlier research (Bongard-Blanchy et al., 2021; Schaffner et al., 2025).

Third, comparing planned time allocations to realized ones, I find that participants consumed significantly less content than their day 1 plans implied in both conditions. The decrease amounts to 35% less time spent on the watching task. This result is driven by 33% of all participants who had planned to spend up to three quarters of their time watching videos but instead decreased their content consumption on the next day. On the other end, 88% of those who had initially planned to spend less than a quarter of their time watching videos successfully followed their plans, matching their planned time allocations.

Fourth, evidence from a regression discontinuity analysis shows that participants monitored their day 1 plans despite ultimately exceeding them. When participants reached their planned transcribing duration, they immediately reduced transcribing probability by 25.5 percentage points in the next 60-second window. However, this effect dissipated quickly and became statistically insignificant within 90 seconds as participants resumed transcribing.

Finally, a mixed-methods analysis of open-ended responses reveals that participants perceived the experimental environment as a work setting rather than a choice between independent tasks. While 29% spontaneously described transcribing using negative terms, another 23% characterized video-watching as "taking breaks" from work. This suggests that experimenter demand effects may have confounded the intended effects of the autoplay treatment, with implications for both the interpretation of our null results and the broader generalizability of our study.

A primary internal validity concern is that experimenter demand effects favored allocating more time to the transcribing task than planned. Such effects can emerge through contextual cues and alter participant behavior by drawing attention to variables of interest (Zizzo, 2010). In my experiment, these could include (i) the small earnings difference favoring the transcribing task,[4] (ii) multiple mentions of the quality criteria for the transcribing task to receive earnings, (iii) placement of the transcribing task as the first activity upon starting a session, and (iv) the time allocation slider framing decisions as time spent on the transcription task. Alternative designs could have addressed these concerns by rewarding balanced time allocations more, eliminating earning differences altogether, or displaying both tasks right next to each other.

Another major challenge in isolating the effects of autoplay is the role of content appeal. Higher content appeal would make the watching task more tempting, and create a stronger tradeoff between the two tasks. To address this while keeping the content identical across conditions, I conducted pretests from the same participant pool where the funny animal videos came out as the most popular theme. Participants also rated the curated videos positively (median 7/10) in a post experiment survey. Regardless, the watching task content may not have been sufficiently appealing. Allowing participants to select among a list of themes with manually curated videos could have addressed this issue (Ek & Samahita, 2023).

A final concern in online experiments is ensuring genuine participation equivalent to laboratory conditions. I implemented several measures to address this challenge. I used JavaScript-based attention monitoring to detect each second participants navigated

[4]This difference is 60 pence over 20 minutes, equivalent to a 1.8 pound hourly wage difference (15% above our 12 pound minimum wage baseline). I hypothesized that in the absence of an earnings difference, participants would not work on the transcription task.

away from the experimental interface following (Purohit et al., 2023). I also required minimum internet speeds and restricted participation to desktop users with Chromium-based browsers to standardize the user experience. Participants also answered post-experiment survey questions about their engagement while watching videos and about connection interruptions. These controls ensured that my final dataset is composed of participants who were active, and the online environment could approximate the conditions of a laboratory experiment (Arechar et al., 2018).

This study contributes to several strands of literature. First, the literature studying the effects of social media and streaming services on user well-being (Allcott et al., 2020, 2022; Bao, 2025; Beknazar-Yuzbashev et al., 2025; Braghieri et al., 2022; Bursztyn et al., 2023; Collis & Eggers, 2022; Groshek et al., 2018; Nyhan et al., 2023; Purohit et al., 2023; Walton-Pattison et al., 2018). I contribute to this work by shedding light on one potential driver of the harmful effects of social media and streaming services, the autoplay, in a controlled setting.

A subset of this literature focuses on the effects of specific interface design elements on user behavior, particularly related to user agency and overconsumption of content (Bongard-Blanchy et al., 2021; Chen et al., 2025; Hoong, 2021; Lukoff et al., 2021; Lupiáñez-Villanueva et al., 2022; Lyngs et al., 2019, 2020; Mathur et al., 2021; Schaffner et al., 2023, 2025; Silverman et al., 2024). To my knowledge, this paper provides the first attempt at isolating the causal effects of autoplay on content consumption.

Methodologically, this paper contributes to the literature on laboratory experiments that observe behavior in real time while participants trade cognitive effort with leisure (Bhatia et al., 2021; Bonein & Denant-Boèmont, 2015; Ek & Samahita, 2023; Houser et al., 2018; Kool & Botvinick, 2014). It also adds to a burgeoning literature manipulating design elements within the user interface to reduce overconsumption of content (Hiniker et al., 2018; Lyngs et al., 2020; Purohit et al., 2023; Schaffner et al., 2025).

The remainder of this paper is organized as follows. Section 2 describes the experimental design. Section 3 details the experimental sample and procedures. Section 4 presents results and Section 5 concludes.

2 Design

2.1 Timeline and Session Structure

I conducted a two-day online lab experiment to examine how autoplay feature affects time allocation between transcription of random characters and watching funny animal videos. The multi-day structure allowed us to measure the gap between planned and actual behavior, a key feature in research on present-focused preferences (Ericson & Laibson, 2019). Participants completed two separate data collection sessions with a cooling-off period of 24 hours between sessions.

Day 1 sessions included a practice session to familiarize participants with the experimental interface, followed by planned time allocation decisions. Day 2 featured the main session where participants made actual time allocation decisions. I conducted two waves of data collection during May 2023, as described in Table 1.

Table 1: Summary of timeline and session structure

	Practice Session	Multiple Price List	Time Choice	Main Session	Decision Horizon
Day 1 (Block 1)	✓		✓		24-hour
Day 2 (Block 1)				✓	immediate
Day 1 (Block 2)	✓	✓	✓		24-hour
Day 2 (Block 2)				✓	immediate

Note: This table shows the experimental components administered across two data collection waves, with columns ordered chronologically within each day. The first block established differences in planned and actual time allocations across conditions, while the second block incorporated the Multiple Price List to measure demand for commitment against autoplay. Decision horizon refers to the period between planned time allocation and the main session.

2.2 Technical Implementation and Interface Design

I designed the experimental platform to provide precise control over the testing environment. I presented both tasks within a single web page using distinct tabs, allowing participants to switch between transcribing and watching videos seamlessly while preserving progress. A progress bar and timer provided feedback for remaining time. Sessions began with the transcription task as the default, requiring an active choice to switch to the watching task.

The interface incorporated several key features to track behavior. I logged all critical interactions including tab switches, keystrokes during transcription, video controls, and cursor movements outside the experimental window. This granular tracking enabled me to construct detailed measures of attention and engagement across tasks.

2.3 Internal Validity Controls

Online experiments present unique challenges for maintaining experimental control, including participant multitasking, technical disruptions, and environmental variation. I implemented comprehensive measures to address these concerns and ensure data quality. I clearly communicated these restrictions to participants at the experiment's outset.

Participant and Technology Screening: I selectively admitted participants using multiple screening criteria to ensure a stable experience. Internet speed requirements filtered out participants with unstable connections (minimum 30 Mbps for block 1, 40 for block 2). I restricted participation to users with Windows, Linux, or Mac operating systems using Chromium-based browsers exclusively. I excluded Safari and Mozilla Firefox due to restrictive autoplay configurations. Mobile devices and tablets were prohibited to standardize the user experience.

Platform Controls: The experimental website enforced linear navigation, preventing participants from moving backward through the experiment. I disabled right-click functions and keyboard shortcuts. Any attempt to navigate backward reset progress to the landing page.

Attention and Engagement Monitoring: I implemented real-time tracking of participant attention using a JavaScript library to detect interface visibility disruptions.[5] This system captured actions such as minimizing the browser, switching between web pages, or launching other applications. The tracking provided second-by-second measures of participant engagement.

Post-Experiment Validation: After completing the experiment, participants reported any connectivity disruptions and confirmed their genuine engagement with the watching task. I combined these self-reports with tracking measures to identify and filter data potentially compromised by distractions or inattention.

2.4 Tasks

2.4.1 Transcription Task

The transcription task requires transcribing successive CAPTCHAs. CAPTCHAs are computer-generated images that contain random letters, numbers and randomly placed white spaces.[6] Each CAPTCHA has a total length of 35 characters, including the white spaces. Participants spent on average 41 seconds per CAPTCHA.

I designed the images to be blurred and distorted by lines and dots added on top to increase comprehension difficulty.[7] The interface always presents these CAPTCHA images with a text box and a submit button. When the submit button is clicked, the next CAPTCHA image appears and the text input for the previous one is stored. The task is identical across conditions, as illustrated in Figure 1.

I established easy-to-meet quality requirements to ensure participants remained active. These requirements include an overall transcription accuracy of at least 70 percent in submitted CAPTCHAs[8] and a minimum of one CAPTCHA submission per minute

[5] See more information about the Intersection Observer API.

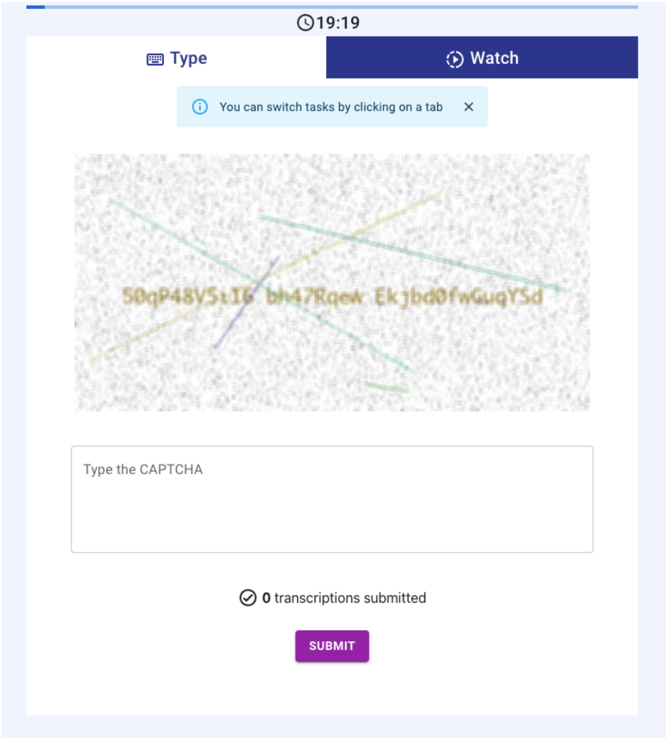
[6] I added white spaces to improve readability after pretests. The number of white spaces was limited to a maximum of 5 instances per CAPTCHA to control the transcribing time per image.

[7] The randomization procedure to generate CAPTCHAs.

[8] I use Python's `difflib` string-matching algorithm to measure accuracy. It consists of finding the longest common substring and then finding recursively the number of matching characters in the non-matching regions on both sides of the longest common substring. Overall accuracy of 70 percent implies

spent on the task. Participants had to meet these quality requirements to receive compensation for the transcription task.

Figure 1: Transcription task tab



Note: This figure describes the transcription interface tab. As participants always begin in the transcription tab by default, there is a banner that reminds participants to click the tab buttons to switch tasks. This banner disappeared as soon as participants switched tasks once.

2.4.2 Watching Task

In this task, participants watch successive short videos in a customized media player. The video content and the order in which videos appear to participants are identical across conditions. I completely disabled the media player controls: participants cannot skip videos (user scrolling is disabled) or use the media player's slider to advance or go back in a particular video.

The conditions differ in their autoplay functionality. In the *Autoplay* condition,

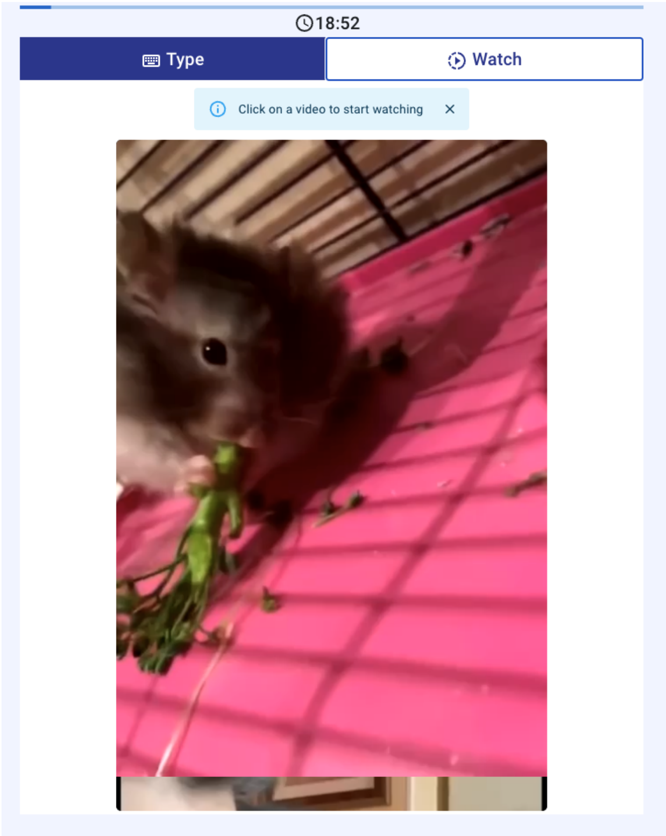
an average of 10.5 $(0.3 * 35)$ mistakes across all submitted CAPTCHAs.

videos play automatically when the watching tab is visible and pause with a click anywhere on the video. The media player keeps scrolling to the next videos and continues playing them unless interrupted. In the *Control* condition, participants need to have the watching tab visible and click on each consecutive video to play them. At the end of each video, the scrolling animation brings the next video that only starts playing when the participant clicks on it. Figure 2 illustrates the video tab in the *Control* condition.

I manually selected 80 videos to avoid algorithmic biases resulting from the recommender system's personalized suggestions. I curated the videos from YouTube and TikTok, using the tags "funny animal" and "cute animal". I chose animal videos because of their popularity during pretests.[9] I reviewed the selected videos to ensure that they did not contain any harmful or violent actions toward the animals involved. To maximize the frequency of experiencing the autoplay feature during the watching task, I kept the first 60 videos shorter (mean 12.25 seconds, SD 7.04 seconds), with the last 20 videos longer than the rest (mean 25.50 seconds, SD 14.04 seconds).

[9]During the pretests, the experiment was followed by a separate questionnaire asking participants to rate the videos they had seen and suggest what types of videos they would have liked to see more of, with options including sports, news, or trending content.

Figure 2: Watching Task tab



Note: This figure describes the "Watch" tab in the *Control* condition. Before the first video is clicked upon to be played, there is a banner that reminds participants that the video won't play without clicking. This banner was not present in the *Autoplay* condition and videos started playing as soon as participants clicked on the "Watch" tab seen on the top right side. Notice that the top part of the following video is already visible in the bottom part of the screen.

2.5 Practice Session

The experiment consists of two sessions during which participants interact with the tasks described above. The first session is the practice session on day one. It comes right after the written descriptions of each task and serves to familiarize participants with the tasks and the interface. Participants are assigned to either *Autoplay* or *Control* condition before the practice session, so that they experience the same video interface on both the practice and main sessions.

Practice session was 2 minutes long. Each task was available for only 60 seconds and participants were not allowed to switch between tasks. On the first tab, participants were required to retype CAPTCHAs. The 60-second period was enough to submit one CAPTCHA and see the second CAPTCHA image. After 60 seconds, a pop-up appeared on the screen blocking any other interaction and taking participants to the second page.

Participants were required to close this pop-up to advance to the second task. The rationale behind the pop-up was to ensure participants spent equal time on both tasks. Whether this pop-up was closed was verified and served as an additional attention check.[10] Once closed, the watching task began for 60 seconds, enabling participants to watch 4 to 5 short videos depending on whether they were assigned to the *Autoplay* or the *Control* condition.

2.6 Multiple Price List

In the second block of the experiment, I used a Multiple Price List to elicit the Willingness to Pay for commitment against autoplay, commonplace method in the literature (Andersen et al., 2006; Jack et al., 2022). After the practice session on day one, participants made 9 binary decisions, choosing between their preferred media player setting (autoplay off or on) and a bonus payment (see Figure 3). I determined bonuses by the per-second earning differences across tasks, multiplied by the expected differences in time spent between the conditions in the first block of the experiment.

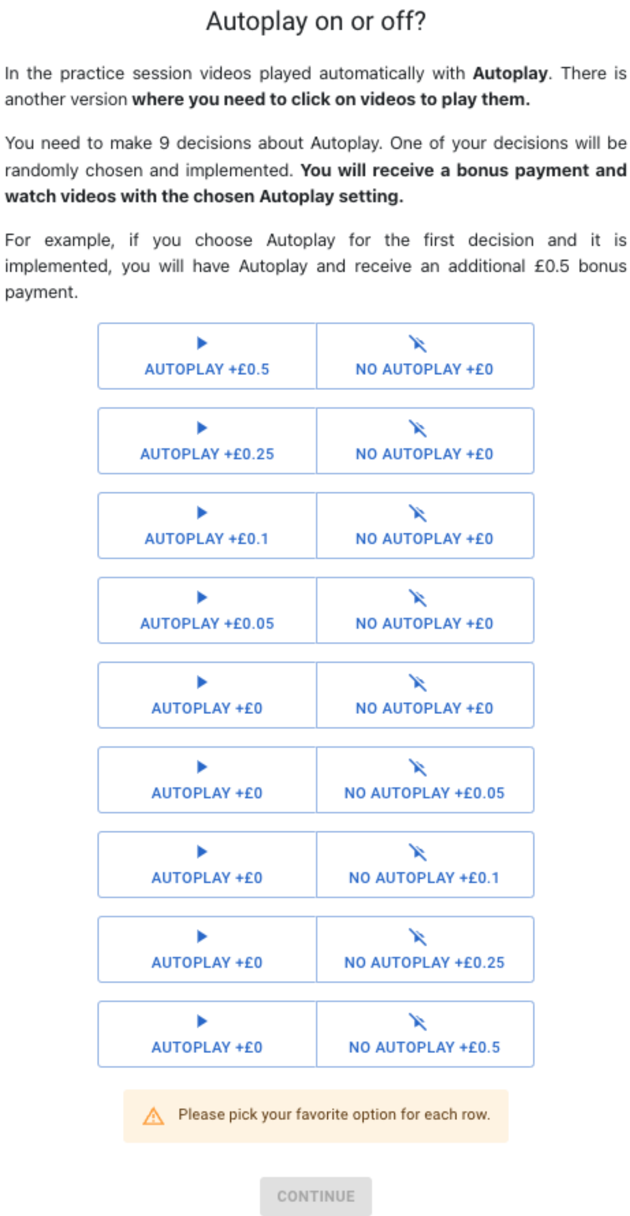
Specifically, I expected the average difference between the *Autoplay* and *Control* conditions, in terms of time spent across tasks, to be in the magnitude of 60 seconds (it was 79.5 seconds, from Table 3). This resulted in an average earning difference of $60 * (0.15 - 0.1) = 3$ pence. To capture this in the MPL choices offered, I set the symmetric bonuses accordingly: £[0.05, 0.1, 0.25, 0.5].

I used the Becker-deGroot-Marschak (Becker et al., 1964) random price mechanism for incentivization. Once the binary decisions were completed, I randomly implemented

[10] All participants closed this pop-up and switched to the watching task.

one of their 9 choices and informed participants whether autoplay would be on or off, and of their associated bonus earnings. With this knowledge, participants then made their planned time allocation for day 2.

Figure 3: Multiple Price List page



Note: This figure describes the MPL page where participants made 9 binary choices between their preferred autoplay setting and bonus payments. Each row presents a choice between "Autoplay" (left) and "No Autoplay" (right) with associated bonus amounts ranging from £0.05 to £0.5. Note that participants in this block only saw autoplay videos in the practice session, and stated their preferences accordingly. The interface required participants to select one option in each row before proceeding. One of these 9 decisions was randomly selected for implementation in the main session, determining both the autoplay setting and any bonus payment for day two.

2.7 Time Choice

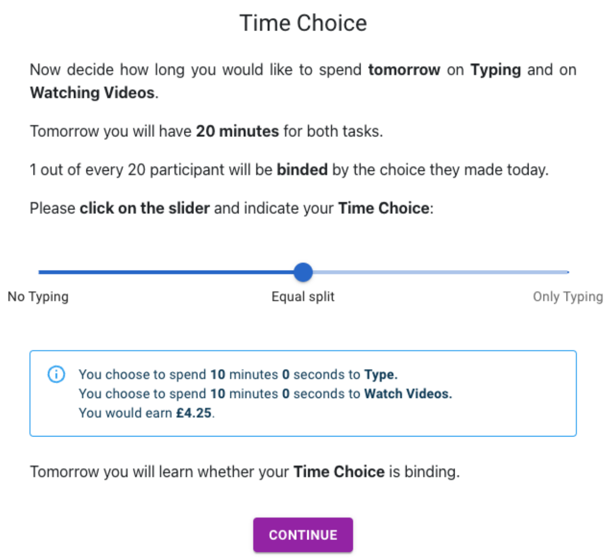
Following the practice session, participants proceeded with the time choice section. This section required them to specify their preferred time allocation between the two tasks for the day 2. Participants were informed the main session would be 20 minutes long, and they were free to choose corner solutions, even if that meant focusing solely on one task.

I designed a slider tool with 20-second steps for participants to indicate their time allocation preferences. Initially, the slider appeared grayed out and required a click to activate. This design aimed to ensure participants considered their choices in an active manner. The click-to-activate design had two benefits: It guaranteed participants interacted with the slider and helped participants understand the task-related payoff structure.

To incentivize this decision, I randomly made 5% of the day 1 choices binding in the main session. Binding choices led to payments based on the initial time choice, but only if participants met the quality standards described under Section 2.4.1 during day 2. I excluded data from these participants from the final dataset accordingly.

The time choice served as a benchmark for participants, and participants accordingly received a reminder of it on day two. The time choice section offers insights into participants' informed time preferences and allows verifying their commitment to these choices after a 24-hour period.

Figure 4: Time Choice page



Note: This figure shows the time choice interface with a slider tool with 20-second steps allowing participants to allocate time between the two experimental tasks for the following day's main session. The slider initially appears grayed out and requires a click to activate. Real-time feedback shows the selected allocation (here, 10 minutes each) and expected earnings (£4.25). The interface informs participants that 1 out of 20 (5 percent) will be bound by their choice and that they will learn whether their choice is binding on day two.

2.8 Main Session

The main session took place on day two, approximately 24 hours after the time choice. I began by reminding participants of their previous day's time allocation and informing them whether their time choice was binding. 95% of participants were free to allocate their time on day 2 as they preferred.

The main session lasted 20 minutes. Participants could switch between the transcription and watching tasks at any time using the tab interface. Unlike the practice session, there were no time restrictions on individual tasks. Participants had complete autonomy over their time allocation decisions, and could monitor remaining time through the progress bar and the timer. Throughout the main session, I monitored writing and transcribing actions in real-time.

The treatment was implemented through the media player's autoplay functionality as previously described in Section 2.4.2.

2.9 Hypotheses

User interface design choices are "not separable from other content choices, since their effect depends on where the platform has positioned its content on the underlying content manifold." (Kleinberg et al., 2022). If content has very low inherent appeal and stickiness, autoplay may have minimal effect on video consumption. An analogy is how introducing breaks would have little impact on engagement for low-stickiness content like documentaries, but dramatically reduce engagement for high-stickiness content like celebrity gossip.[11] Assuming that my selection of videos have sufficient appeal and stickiness, *Autoplay* condition leads to consumption beyond users' stated preferences, creating a measurable gap between planned the time allocation and actual behavior.

Hypothesis 1: *Autoplay* condition decreases the realized time allocated to the transcription task while increasing video consumption. I test this using a linear regression:

$$Y_i = \alpha + \gamma \text{Autoplay}_i + \beta \text{TimeChoice}_i + \delta(\text{Autoplay}_i \times \text{TimeChoice}_i) + \epsilon_i \qquad (1)$$

where Y_i represents the outcome variables of interest (proportion of time spent on transcription or number of videos watched), Autoplay_i is a binary indicator for the condition, and TimeChoice_i is the standardized proportion of time participants planned to spend on transcription on day 1. I estimate this model with and without the interaction term to examine whether the effect of autoplay varies based on initial time allocation plans. I expect $\gamma < 0$ for transcription and $\gamma > 0$ for videos watched.

Hypothesis 2: *Autoplay* condition has an effect on deviations from stated time allocation preferences (actual minus planned time spent on transcription). I test this using a two-sample t-test comparing the means between the two conditions.

[11] See Bao (2025) for recent work on addictive short drama series.

I expect the effect of autoplay on content consumption is known by participants, although they will underestimate how much they would be affected on average (Bongard-Blanchy et al., 2021). The demand for a commitment device is driven by "sophisticated" participants, i.e., a subset of participants who are aware of the effect of autoplay, in line with the literature (Ericson & Laibson, 2019).

Hypothesis 3: If participants are aware of their self-control issues, they should exhibit negative Willingness To Pay (WTP) for autoplay, preferring the commitment devices that restrict automatic video transitions. I test this using a logistic regression:

$$Y_{ij} = \beta_0 + \beta_1 \text{Bonus}_{ij} + \epsilon_{ij} \tag{2}$$

where Y_{ij} is the binary choice of taking autoplay given the bonus j from the MPL for individual i. I calculate the WTP for autoplay as the sample-averaged effect of bonus $-\beta_0/\beta_1$, and expect a negative sign indicating preference for commitment against autoplay.

In sum, the preregistered hypotheses test whether autoplay distorts optimal time allocation decisions in a setting where participants face monetary incentives favoring the transcription task.

3 Sample and Procedures

The experiment was run entirely online, using widely available open-source tools.[12] I sampled from the Prolific participant pool, an online recruitment platform commonly used for academic research, and targeted adult British residents. I further imposed the following criteria: Fluency in English, balanced gender representation, and having a stable internet connection. I did not allow mobile devices and restricted access to the experiment to Chromium-based browsers to enhance the internal validity as previously discussed in Section 2.3. Only participants who succeeded on both attention checks, reported not having engaged in other activities, and did not have connection issues during the experiment were included in the final dataset.

[12] I used the MERN stack to develop and host the website in a free way.

Payment was calculated based on actual time spent on each task during the main session on day 2, with compensation rates of 0.15 pence per second for transcription and 0.1 pence per second for watching. Participants needed to meet the established quality requirements for the transcription task to receive compensation for time spent on that activity. The session concluded automatically after 20 minutes, and participants received their final payment calculation before completing the experiment. Participants received a fixed participation fee of £2.75. The study took less than 30 minutes in total across 2 days, with average earning per participant at £4.08. This amounted to an average payment of £1.33 (median £1.49) for the main session. Participants who completed all required elements received payment within seven days after the second session.[13]

For the first block of the experiment, I preregistered a two-sample parallel design using pretest data and tested for equality of means across conditions. I found that the required sample size to test Hypothesis 1 with the usual parameter values ($\alpha = .05$, $\beta = .2$) would be 114 per condition, with an effect size of 65 seconds calculated by the difference in the average time spent between conditions ($\mu_{Autoplay} - \mu_{Control}$) and population standard deviation of 165 seconds. Since attrition in longitudinal experiments is commonplace, I invited 301 participants on the first day. On day two, I invited the same 301 participants and received 276 complete responses (91% response rate). No participants failed the attention checks more than twice.[14] I dropped 17 participants who self-reported having connection issues and 24 additional participants who stated having engaged in other activities. I also removed 11 participants who had their time choice enforced randomly. Finally, I removed 40 participants that were detected to spend more than 20 consecutive seconds away from our experimental platform who disproportionately come from the *Autoplay* condition. These 40 participants all reported engaging with videos but had their mouse detected outside the experiment window, making it impossible to distinguish whether they genuinely watched videos or browsed other tabs/windows during the study (see Appendix Table A.1). The final dataset on

[13] Prolific payment policy gives researchers 21 days to transfer funds to participants. The typical delay between study completion and payment is however only 3 to 4 days on average.

[14] Minimum requirement by Prolific attention check policy.

which I base the analysis consisted of 184 individuals. Table 2 shows the demographic characteristics of the observed individuals.

Table 2: Balance table for the sample

	Autoplay N = 79	Control N = 105	p [A = C]
Gender			
Female	36 (45.57%)	53 (50.48%)	0.510
Age			
Mean (SD)	42.63 (11.70)	42.03 (11.24)	0.725
Employment			
Employed (full time)	45 (56.96%)	57 (54.29%)	0.718
Employed (part time/self)	23 (29.11%)	25 (23.81%)	0.417
Not employed	11 (13.92%)	23 (21.90%)	0.167
Income			
Low (0-15)	24 (30.38%)	32 (30.48%)	0.989
Middle (15-50)	42 (53.16%)	61 (58.10%)	0.505
High (50+)	13 (16.46%)	12 (11.43%)	0.325
Marital Status			
Married/Partnership	43 (54.43%)	52 (49.52%)	0.510
Single	31 (39.24%)	46 (43.81%)	0.534
Previously married	5 (6.33%)	7 (6.67%)	0.927

Note: This table presents demographic characteristics for the sample of 184 participants across conditions. Income categories are presented in thousands of British pounds. Employment categories: "Not employed" includes unemployed (looking and not looking) and homemaker. "Previously married" includes divorced, widowed, and separated. p-values for binary outcomes are from two-sample tests of proportions; for continuous variables, from two-sample t-tests. For employment, part-time and self-employed are together as one category. The sample demonstrates successful randomization with balanced participant characteristics across the conditions.

For the second block of the experiment, I applied the same selection criteria as before and invited 60 participants on day one. On day two, I received 57 complete answers (95% response rate). Once again, no participant failed the attention checks. No participant reported having connection issues. I dropped 5 participants who reported having engaged in other activities. As I was interested in the demand for a commitment device, participants did not have their Time Choice enforced in this treatment. I therefore ended up with 52 individuals in the final MPL dataset (see Appendix Table A.2 for a comparison with the main sample).

4 Results

4.1 Does Autoplay increase video consumption?

I test my primary hypothesis that autoplay increases video consumption by comparing behavior across conditions. Table 3 presents summary statistics for key outcome variables from the 20-minute main session. Two-sample t-tests reveal no significant differences between the *Autoplay* and *Control* conditions.

Table 3: Sample statistics for variables of interest

	Autoplay N = 79	Control N = 105	p [A = C]
Time Choice			
Mean (SD)	0.48 (0.31)	0.44 (0.33)	0.410
Median [Min, Max]	0.50 [0.00, 1.00]	0.50 [0.00, 1.00]	
Transcription Proportion			
Mean (SD)	0.65 (0.31)	0.58 (0.33)	0.173
Median [Min, Max]	0.72 [0.00, 1.00]	0.56 [0.00, 1.00]	
Nb. of sessions			
Mean (SD)	5.76 (4.98)	4.89 (5.09)	0.246
Median [Min, Max]	4.00 [1.00, 18.00]	3.00 [1.00, 28.00]	
Session length			
Mean (SD)	456.11 (383.22)	546.35 (411.53)	0.131
Median [Min, Max]	314.75 [68.17, 1212.00]	418.67 [44.93, 1314.00]	
Nb. of CAPTCHA submissions			
Mean (SD)	19.76 (12.11)	17.21 (11.62)	0.150
Median [Min, Max]	19.00 [0.00, 53.00]	15.00 [0.00, 45.00]	
Nb. of videos watched			
Mean (SD)	31.33 (26.34)	32.72 (26.39)	0.723
Median [Min, Max]	25.00 [0.00, 80.00]	34.00 [0.00, 77.00]	
Content rating			
Mean (SD)	6.96 (2.18)	6.62 (2.45)	0.327
Median [Min, Max]	7.00 [2.00, 10.00]	7.00 [0.00, 10.00]	

Note: This table presents key outcome variables from the Main Session. "Time Choice" refers to participants' stated preferences for transcribing on day 1 (proportion), while "Transcription Proportion" represents the actual proportion of time spent on transcription task during the 20-minute Main Session on day 2. Nb. of sessions is the number of transcription or watching sessions participants completed, and Session length represents the mean duration of individual sessions in seconds. Content rating reflects participants' appreciation of video content on a 0-10 scale. p-values are from two-sample t-tests comparing experimental conditions.

To formally test Hypothesis 1, the effects of the *Autoplay* condition on time spent on transcription, I run the regression specified in Equation 1. My analysis reveals that the autoplay manipulation had no statistically significant effect on actual transcription behavior. Table 4 presents results for transcription proportion as the dependent variable. The coefficient on the autoplay ranges from 0.12 to 0.20 across specifications but remains statistically insignificant in all models (p-values > 0.16). Participants' day 1 time allocations emerge as a powerful predictor of day 2 behavior: a one standard deviation increase in planned transcription time is associated with a 0.67 standard deviation increase in actual transcription time ($p < 0.001$). Including day 1 time allocations increases the variance explained by the model, rising from 1% to 46%.

Table 4: *Autoplay* condition and time spent transcribing

	Transcription proportion (z-score)		
	(1)	(2)	(3)
Autoplay	0.203	0.121	0.119
	(0.147)	(0.109)	(0.110)
Time Choice (z-score)		0.673***	0.651***
		(0.057)	(0.081)
Autoplay × Time Choice			0.057
			(0.110)
Constant	−0.087	−0.052	−0.053
	(0.100)	(0.076)	(0.076)
Obs.	184	184	184
R^2	0.010	0.462	0.463
F-statistic	1.91	77.42	61.52

Note: Robust standard errors in parentheses. All variables are standardized. Time choice represents participants' planned transcription proportion from day 1. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Similarly, Table 5 examines the effect on video consumption, a direct test of my primary hypothesis that autoplay would increase video consumption. Across all specifications, the autoplay coefficient is statistically insignificant. This suggests that the autoplay feature did not meaningfully alter participants' video consumption behavior. Again, day 1 plans prove to be the dominant predictor: participants who planned more transcribing watched −0.66 SD fewer videos ($p < 0.001$).

Table 5: *Autoplay* condition and the number of videos watched

	Nb. of Videos Watched (z-score)		
	(1)	(2)	(3)
Autoplay	−0.053	0.029	0.031
	(0.149)	(0.112)	(0.112)
Time Choice (z-score)		−0.664***	−0.619***
		(0.056)	(0.079)
Autoplay × Time Choice			−0.114
			(0.105)
Constant	0.023	−0.012	−0.010
	(0.098)	(0.076)	(0.076)
Obs.	184	184	184
R^2	0.001	0.440	0.443
F-statistic	0.13	73.06	61.53

Note: Robust standard errors in parentheses. All variables are standardized. Time choice represents participants' planned transcription proportion from day 1. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

For completeness, I provide the regression outputs from the unrestricted dataset in Appendix Table A.3 and A.4. These include 40 participants who spent at least 20 consecutive seconds outside either task. Comparing the interaction models (column 3), I observe that the effect of the *Autoplay* condition on transcribing behavior decreases from 0.119 SD in the restricted sample to 0.055 SD in the unrestricted sample. The effect of the *Autoplay* condition on the number of videos watched increases from 0.031 SD in the restricted sample to 0.132 SD in the unrestricted sample. However, the null results remain unchanged across both specifications. Because I cannot ascertain the data quality for the unrestricted sample, I base my conclusions on the restricted sample of 184.

Two key patterns emerge that help explain the null findings. First, the contradicting results: the *Autoplay* condition increases time spent transcribing by 0.12 standard deviations (though statistically insignificant) while simultaneously increasing the number of videos watched by 0.03 standard deviations (also insignificant). If people type more, shouldn't they watch fewer videos? I identify a mechanical confound in my experimental design that explains this result: the time lost during video transitions. Participants in the *Control* condition, who had to manually click to start each new video, lost an average of 46.52 seconds due to these transition delays whereas those in the *Autoplay* condition lost only 1.08 seconds. This large difference of 45.45 seconds per session is statistically significant ($t(105) = 4.89$, $p < 0.001$). This difference in time lost not watching videos explains the contradicting results and underscores the importance of looking at the unbiased outcome variable (i.e., number of videos watched) when evaluating the effects of the autoplay feature.

Second, assuming the actual effect size for the number of videos watched lies somewhere between the restricted and unrestricted samples, the power calculations based on pretests were insufficient: The required sample size to provide a two-sided test for Hypothesis 1b with the usual parameter values ($\alpha = .05$, $\beta = .2$) would be between 300 and 5000 per treatment, with common standard deviation as the root MSE from the standardized number of videos watched variable in the third regression. Other solutions to this power problem are design-related and involve extending the main-session duration to above 20 minutes, or nudging participants to mix more between tasks to avoid corner solutions (e.g., higher earnings for time allocations between .25 and .75 of either task).

Table 6: Seconds lost by condition

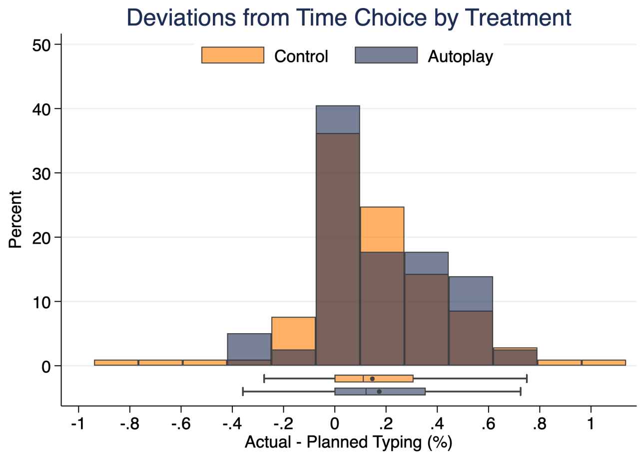
	Control	Autoplay	Difference
Total seconds lost	46.52	1.08	45.45***
	(9.27)	(0.65)	(9.29)
Observations	105	79	184
t-statistic			4.89

Note: Standard errors are in parentheses. Time is measured as the average of seconds lost to video transitions per session. The difference was tested using a two-sample t-test with unequal variances. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

4.2 Autoplay and deviations from day 1 time allocation

Does the *Autoplay* condition have an effect on deviations from day 1 time choice? To assess Hypothesis 2, I calculated the deviation by subtracting participants' planned transcribing proportion from their actual transcribing proportion (actual minus planned). Figure 5 shows the distribution of deviations from time choice across conditions, which appears right-skewed with positive values indicating that participants transcribed more than their day 1 allocation. Participants in the *Control* condition deviated by an average of 14.6 percentage points, while those in the *Autoplay* condition deviated by 17.3 percentage points. A two-sample t-test reveals no statistically significant difference between means ($t = -0.694$, $p = 0.488$). The Kolmogorov-Smirnov test similarly finds no significant difference in distributions across conditions ($D = 0.085$, $p = 0.903$). These results indicate that participants in both conditions spent more time on the transcription task than they initially planned. However, the *Autoplay* condition did not significantly alter the magnitude or pattern of these deviations. Contrary to my hypothesis, I find no evidence that *Autoplay* condition led participants to deviate toward consuming more videos.

Figure 5: Distribution of deviations from day 1 time allocation by condition



Note: This figure shows the distribution of deviations from day 1 time allocation plans, calculated as actual transcribing proportion minus planned transcribing proportion. Positive values indicate participants transcribed more than they planned. The distributions are similar across conditions, and show over-transcribing relative to initial plans.

4.3 Willingness To Pay for autoplay

Contrary to Hypothesis 3, participants exhibited positive WTP for autoplay rather than demand for commitment devices against it. I assessed the demand for commitment by comparing the share of participants choosing to turn autoplay on or off along with the associated bonus payments for each of the nine decisions in the price list. To quantify participants' WTP for autoplay, I estimate the regression specification in Equation 2, where the dependent variable is the binary choice of selecting autoplay (1) or turning it off (0), and the independent variable is the bonus amount in pounds from the MPL.

Results are presented in Table 7. The coefficient on the bonus amount is 27.97 ($p < 0.001$), indicating that participants are sensitive to the financial cost of keeping autoplay on. The WTP for autoplay is the negative ratio of the constant to the bonus coefficient: $-0.6263/27.9729 = -0.0224$ pounds, or 2.24 pence for the 20-minute session. On average, participants value autoplay at 6.72 pence/hour and would be willing to forgo this amount to maintain access to the feature rather than have it turned off. Participants in this study did not appear to perceive a tradeoff from autoplay in increasing video consumption.

Table 7: Willingness to Pay for autoplay

	(1)
Bonus (£)	27.973***
	(6.277)
Constant	0.626**
	(0.199)
Obs.	468
Clusters	52
Pseudo R^2	0.658
Wald $\chi^2(1)$	19.86

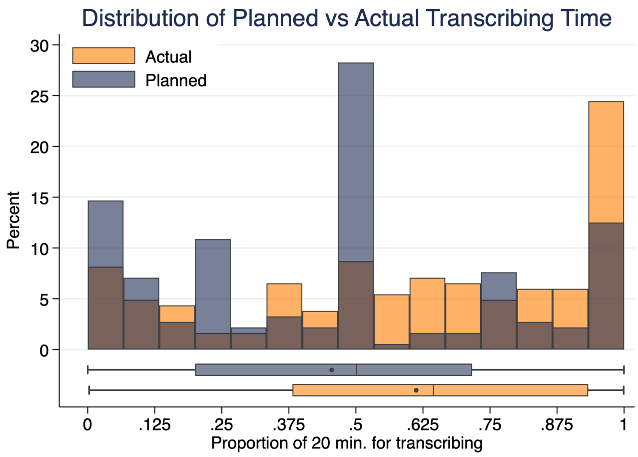
Note: Logistic regression with standard errors clustered at the participant level in parentheses. Dependent variable is binary choice of selecting autoplay (1) or turning it off (0). Bonus amounts range from £−0.5 to £0.5. The pseudo R^2 indicates good fit. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

4.4 Understanding the Null Results

4.4.1 Planned versus Actual Time Allocation

Participants allocated more time to transcription than initially planned. On average, participants planned to spend 45.5% of their time transcribing but actually spent 61.2%, representing an increase of 15.7 percentage points ($t = 8.24$, $p < 0.001$). The entire distribution shifted rightward (see Figure 6), with median actual transcription time (64.3%) substantially exceeding median planned time (50.0%).

Figure 6: Distribution of planned vs actual time allocation



Note: This figure compares the distribution of participants' planned time allocation (Day 1) with their actual time spent transcribing (Day 2). The rightward shift demonstrates participants' systematic tendency to transcribe more than originally planned.

The pattern is more pronounced when examining individual transitions. Among 151 participants who planned to spend less than three-quarters of their time transcribing, 106 (70%) actually transcribed for three-quarters or more of the session. In contrast, among 33 participants who planned to spend three-quarters or more of their time transcribing, only 4 (12%) reduced below 75%. This asymmetric upward shift contradicts traditional models of time inconsistency, which predict succumbing to immediate temptation rather than increasing effortful work.

Table 8: Transition matrix: Planned vs actual time allocation

	Planned (Day 1)	
Actual (Day 2)	<75% transcription	≥75% transcription
<75% transcription	45 (30%)	29 (88%)
≥75% transcription	106 (70%)	4 (12%)
Total	151	33

Note: This table shows transitions between planned and actual time allocation patterns. The asymmetric pattern demonstrates systematic shift toward transcription, with 106 participants increasing transcription time versus only 4 decreasing it.

The deviation toward the more effortful transcription task is rather unexpected, and participant perceptions of the decision-making environment may have influenced their behavior. This result goes against my hypothesis that participants would deviate toward the tempting video content, and points to the need to examine it more closely.

4.4.2 Experimenter demand effects

Why did participants end up transcribing more than their day 1 plans? To understand how participants perceived the decision-making environment, I asked how they decided to allocate their time during the main session. I analyzed participants' qualitative responses to this open-ended question using dictionary-based concept detection methods (Ash & Hansen, 2023).

I identified words related to negative perceptions of the transcription task (e.g., "tiresome", "frustrating", "difficult") and break-taking behavior when switching to watching (e.g., "break", "rest", "relax"). Figures 7a and 7b show the frequency of these concepts in participants' responses.

Figure 7: Word clouds for participant perceptions

(a) Typing task negative descriptions

(b) Changing to watching task

Note: Panel (a) shows words describing negative perceptions of the transcription task. Panel (b) shows words describing switching to the watching task as break-taking behavior. Word size corresponds to frequency of mention across responses.

When asked to describe their allocation decisions, 29% of participants described the transcription task negatively, while 23% specifically described watching videos as taking breaks from transcription. These findings suggest that participants perceived the experiment as a work environment where they should transcribe as much as possible while taking breaks to watch the videos. Combined with the systematic increases in transcription time, experimenter demand effects have influenced participant behavior as suspected.

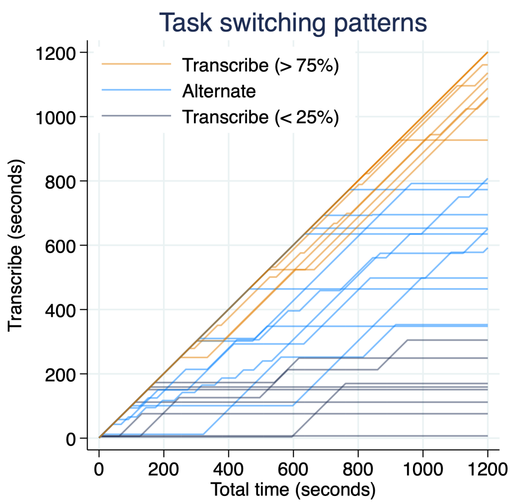
4.4.3 Effect of achieving day 1 time allocation

If experimenter demand effects are driving participants to transcribe more than planned, participants may ignore their day 1 goals entirely and continue transcribing regardless of achieving their stated preferences. Alternatively, if participants genuinely care about their day 1 plans, we should observe behavioral changes when they reach their planned transcribing duration—specifically, a reduction in transcribing effort after achieving their goal.

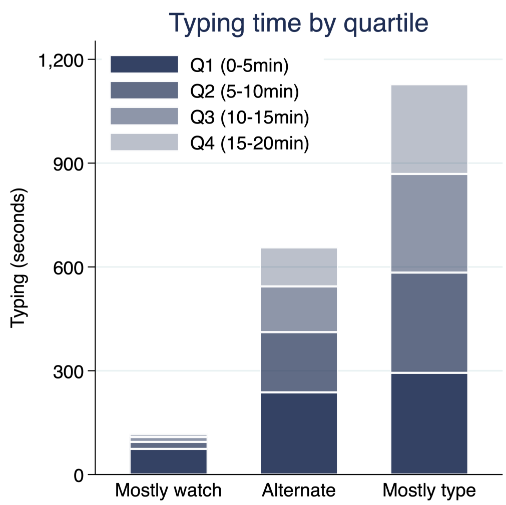
I examine this by analyzing task switching patterns throughout the 20-minute session. Figure 8a shows substantial heterogeneity in individual transcribing patterns, with transcribing periods interrupted by watching periods for participants who alternate between tasks. Dividing the session into 5-minute quarters reveals that transcribing periods become progressively shorter over time (Figure 8b), suggesting that task switching may be strategic as participants either achieve their day 1 goals or experience fatigue.

Figure 8: Time spent transcribing

(a) Individual transcribing patterns



(b) Average transcribing by quartile



Note: Panel (a) shows transcribing time for 30 random participants across the 20-minute session. Panel (b) shows average transcribing time by 5-minute periods, revealing declining transcribing duration over time.

To test whether participants use their day 1 plans as behavioral reference points, I implement a sharp regression discontinuity design using the moment participants reach their planned transcribing duration as the cutoff. I estimate the discontinuous change in the share of participants choosing to transcribe task at this threshold.

Results in Table 9 show a significant effect: participants are 27.4 percentage points less likely to continue transcribing in the 30-second window after achieving their plan

($p = 0.007$) and 25.5 percentage points less likely in the 60-second window ($p = 0.003$). This effect dissipates by the 90-second window, becoming statistically insignificant. These findings indicate that participants do treat plan completion as a behavioral milestone, suggesting that experimenter demand effects were not so strong as to completely override their day 1 preferences.

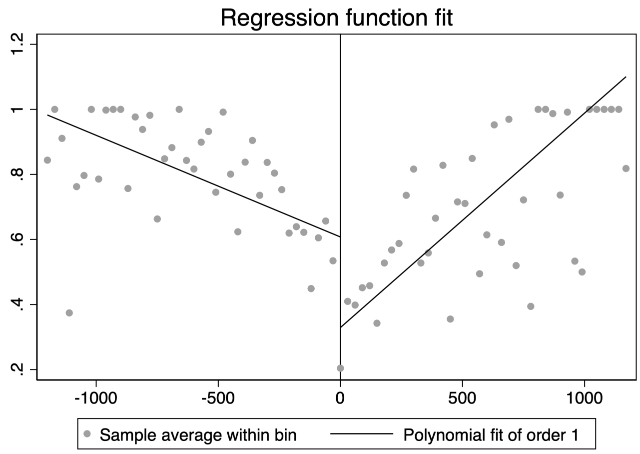
Table 9: Regression Discontinuity estimates on transcribing after achieving day 1 plans

	Time Window		
	30-seconds	60-seconds	90-seconds
Proportion transcribing	-0.274***	-0.255***	-0.075
	(0.101)	(0.087)	(0.108)
Obs. (left)	66,786	67,985	67,119
Obs. (right)	80,270	89,006	92,999
Clusters (left)	157	154	155
Clusters (right)	154	157	161
Bandwidth	339.9	368.3	422.3

Note: This table presents robust regression discontinuity estimates of the treatment effect on task choice at the moment participants achieve their day 1 planned transcribing duration. Standard errors clustered by participant in parentheses. The outcome variable is a binary indicator for choosing the transcribing task. Running variables are binned at 30-, 60-, and 90-second intervals around the achievement threshold. All estimates use triangular kernel with MSE-optimal bandwidth selection. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

However, the discontinuity analysis reveals additional complexity. Figure 9 shows that while participants initially reduce transcribing after achieving their day 1 plans, they subsequently increase transcribing effort again. This pattern suggests that participants have indeed found the watching task less engaging than anticipated, leading them to return to transcribing despite having already met their stated preferences.

Figure 9: Linear fit for transcribing rate before and after achieving day 1 plan



Note: This figure plots the average transcribing rate across 30-second bins for participants as they approach their day 1 goal. The negative slope on the left-hand side indicates participants working less and less as they approach the cutoff. The drop showcases the discontinuity where achieving the day 1 goal causes in transcribing behavior. The positive slope on the right-hand side reveals another trend: As time goes on, participants tend to go back to the transcribing task and decrease their video consumption. This demonstrates participants' tendency to spend more time on the effortful transcribing task than originally planned.

5 Conclusion

This study examined whether autoplay features in isolation can override users' stated preferences for media consumption. In a two-day online experiment with 236 participants, we find no evidence that autoplay increases video consumption when content is held constant. Participants viewed autoplay favorably, exhibiting positive willingness-to-pay for the convenience it provides rather than perceiving it as a self-control problem. However, experimenter demand effects resulted in participants allocating more time to transcription than planned, confounding the intended autoplay treatment.

These findings highlight a fundamental challenge of studying potentially addictive design features in controlled laboratory settings. Design features like autoplay appear to derive their persuasive power from personalized content streams and algorithmic curation. In my controlled setting with identical, manually curated videos, I likely stripped away the very elements that make autoplay compelling in practice.

These results have implications for both research methodology and policy. Laboratory studies that isolate interface features from their natural digital ecosystem may fail to capture their real-world behavioral consequences. The controlled conditions necessary for internal validity inadvertently eliminate the personalized content that drives engagement on actual platforms. While autoplay shows no effect in my experimental setting, field studies suggest stronger impacts in natural usage contexts (Hiniker et al., 2018; Schaffner et al., 2025). Future research should prioritize better approximating genuine media consumption environments in the field (see e.g., Beknazar-Yuzbashev et al. (2025); Purohit et al. (2023)) while maintaining the control necessary for causal inference.

References

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–76.

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

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006, December). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383–405. doi: 10.1007/s10683-006-7055-6

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, *21*(1), 99–131. doi: 10.1007/s10683-017-9527-2

Ash, E., & Hansen, S. (2023). Text Algorithms in Economics. *Annual Review of Economics*, *15*(Volume 15, 2023), 659–688. doi: 10.1146/annurev-economics-082222-074352

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

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232. doi: 10.1002/bs.3830090304

Beknazar-Yuzbashev, G., Jiménez-Durán, R., McCrosky, J., & Stalinski, M. (2025). Toxic content and user engagement on social media: Evidence from a field experiment.

Bhatia, S., Crawford, M. M., McDonald, R. L., Moreno, M. A., & Read, D. (2021). *Inconsistent Planning and the Allocation of Tasks Over Time* (OSF Preprints No. b4mg7). Center for Open Science.

Bonein, A., & Denant-Boèmont, L. (2015). Self-control, commitment and peer pressure: A laboratory experiment. *Experimental Economics*, *18*(4), 543–568. doi: 10.1007/s10683-014-9419-7

Bongard-Blanchy, K., Rossi, A., Rivas, S., Doublet, S., Koenig, V., & Lenzini, G. (2021). " i am definitely manipulated, even when i am aware of it. it's ridiculous!"-dark patterns from the end-user perspective. In *Designing interactive systems conference 2021* (pp. 763–776).

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

Bursztyn, L., Handel, B. R., Jimenez, R., & Roth, C. (2023). *When Product Markets Become Collective Traps: The Case of Social Media* (Working Paper No. 31771). National Bureau of Economic Research. doi: 10.3386/w31771

Chen, Y., Fu, Y., Chen, Z., Radesky, J., & Hiniker, A. (2025). *The Engagement-Prolonging Designs Teens Encounter on Very Large Online Platforms* (No. arXiv:2411.12083). arXiv. doi: 10.48550/arXiv.2411.12083

Collis, A., & Eggers, F. (2022). Effects of restricting social media usage on wellbeing and performance: A randomized control trial among students. *PloS one*, *17*(8), e0272416.

Ek, C., & Samahita, M. (2023). Too much commitment? an online experiment with tempting youtube content. *Journal of Economic Behavior & Organization*, *208*, 21–38.

Ericson, K. M., & Laibson, D. (2019). Intertemporal choice. In *Handbook of Behavioral Economics: Applications and Foundations 1* (Vol. 2, pp. 1–67). Elsevier. doi: 10.1016/bs.hesbe.2018.12.001

Groshek, J., Krongard, S., & Zhang, Y. (2018). Netflix and Ill? Emotional and Health Implications of Binge Watching Streaming TV. In *Proceedings of the 9th International Conference on Social Media and Society* (pp. 296–300). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3217804.3217932

Hiniker, A., Heung, S. S., Hong, S., & Kientz, J. A. (2018). Coco's videos: an empirical investigation of video-player design features and children's media use. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1–13).

Hoong, R. (2021). Self control and smartphone use: An experimental study of soft commitment devices. *European Economic Review*, *140*, 103924. doi: 10.1016/j.euroecorev.2021.103924

Houser, D., Schunk, D., Winter, J., & Xiao, E. (2018). Temptation and commitment in the laboratory. *Games and Economic Behavior*, *107*, 329–344. doi: 10.1016/j.geb.2017.10.025

Jack, B. K., McDermott, K., & Sautmann, A. (2022, November). Multiple price lists for willingness to pay elicitation. *Journal of Development Economics*, *159*, 102977. doi: 10.1016/j.jdeveco.2022.102977

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2022). The challenge of understanding what users want: Inconsistent preferences and engagement optimization. *arXiv preprint arXiv:2202.11776*.

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of experimental psychology. General*, *143*(1), 131.

Lukoff, K., Lyngs, U., Zade, H., Liao, J. V., Choi, J., Fan, K., ... Hiniker, A. (2021). How the design of youtube influences user sense of agency. In *Proceedings of the 2021 chi conference on human factors in computing systems* (pp. 1–17).

Lupiáñez-Villanueva, W. F., Boluda, A., Bogliacino, F., & Liva, G. (2022). Behavioural study on unfair commercial practices in the digital environment: Dark patterns and manipulative personalisation. , 303.

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., ... Shadbolt, N. (2019). Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3290605.3300361

Lyngs, U., Lukoff, K., Slovak, P., Seymour, W., Webb, H., Jirotka, M., ... Shadbolt, N. (2020). 'I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3313831.3376672

Mathur, A., Kshirsagar, M., & Mayer, J. (2021). What Makes a Dark Pattern... Dark?: Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). Yokohama Japan: ACM. doi: 10.1145/3411764.3445610

Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., ... others (2023). Like-minded sources on facebook are prevalent but not polarizing. *Nature*, *620*(7972), 137–144.

Purohit, A. K., Bergram, K., Barclay, L., Bezençon, V., & Holzer, A. (2023). Starving the newsfeed for social media detox: Effects of strict and self-regulated facebook newsfeed diets. In *Proceedings of the 2023 chi conference on human factors in computing systems* (pp. 1–16).

Schaffner, B., Stefanescu, A., Campili, O., & Chetty, M. (2023). Don't Let Netflix Drive the Bus: User's Sense of Agency Over Time and Content Choice on Netflix. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW1), 1–32. doi: 10.1145/3579604

Schaffner, B., Ulloa, Y., Sahni, R., Li, J., Cohen, A. K., Messier, N., ... Chetty, M. (2025). An Experimental Study Of Netflix Use and the Effects of Autoplay on

Watching Behaviors. *Proc. ACM Hum.-Comput. Interact.*, *9*(2), CSCW030:1–CSCW030:22. doi: 10.1145/3710928

Silverman, J., Srna, S., & Etkin, J. (2024). *Can Time Limits Increase Time Spent?* (SSRN Scholarly Paper No. 4381779). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.4381779

Walton-Pattison, E., Dombrowski, S. U., & Presseau, J. (2018). 'Just one more episode': Frequency and theoretical correlates of television binge watching. *Journal of Health Psychology*, *23*(1), 17–24. doi: 10.1177/1359105316643379

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. doi: 10.1007/s10683-009-9230-z

A Additional Figures and Tables

Additional Figures

Additional Tables

Table A.1: Comparison of main sample vs flagged participants

	Main Sample N = 184	Flagged N = 40	p [M = F]
Gender			
Female	89 (48.37%)	13 (32.50%)	0.068*
Age			
Mean (SD)	42.29 (11.41)	42.85 (10.75)	0.768
Employment			
Employed (full time)	102 (55.43%)	27 (67.50%)	0.162
Employed (part time/self)	48 (26.09%)	8 (20.00%)	0.420
Not employed	34 (18.48%)	5 (12.50%)	0.366
Income			
Low (0-15)	56 (30.43%)	6 (15.00%)	0.048**
Middle (15-50)	103 (55.98%)	31 (77.50%)	0.012**
High (50+)	25 (13.59%)	3 (7.50%)	0.291
Marital Status			
Married/Partnership	95 (51.63%)	21 (52.50%)	0.921
Single	77 (41.85%)	17 (42.50%)	0.940
Previously married	12 (6.52%)	2 (5.00%)	0.719
Treatment Assignment			
Autoplay condition	79 (42.93%)	25 (62.50%)	0.025**

Note: This table compares demographic characteristics between the main sample (N=184) and participants flagged for potential inattention. Flagged participants reported engaging with videos but had mouse activity detected outside the experiment window, making it impossible to verify genuine engagement. p-values for binary outcomes are from two-sample tests of proportions; for continuous variables, from two-sample t-tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Comparison of MPL sample vs main sample

	Main Sample N = 184	MPL N = 52	p [M = MPL]
Gender			
Female	89 (48.37%)	26 (50.00%)	0.835
Age			
Mean (SD)	42.29 (11.41)	41.90 (14.14)	0.858
Employment			
Employed (full time)	102 (55.43%)	32 (61.54%)	0.433
Employed (part time/self)	48 (26.09%)	10 (19.23%)	0.311
Not employed	34 (18.48%)	10 (19.23%)	0.902
Income			
Low (0-15)	56 (30.43%)	15 (28.85%)	0.825
Middle (15-50)	103 (55.98%)	30 (57.69%)	0.826
High (50+)	25 (13.59%)	7 (13.46%)	0.981
Marital Status			
Married/Partnership	95 (51.63%)	28 (53.85%)	0.778
Single	77 (41.85%)	19 (36.54%)	0.491
Previously married	12 (6.52%)	5 (9.62%)	0.446

Note: This table compares the demographic characteristics for the main sample (N=184) and the final MPL sample in block 2 (N=52). We observe no significant differences across the two samples.

Table A.3: *Autoplay* condition and time spent transcribing

	Transcription proportion (z-score)		
	(1)	(2)	(3)
Autoplay	0.136	0.055	0.055
	(0.133)	(0.108)	(0.108)
Time Choice (z-score)		0.595***	0.580***
		(0.061)	(0.089)
Autoplay $\times$ Time Choice			0.034
			(0.119)
Constant	−0.063	−0.026	−0.027
	(0.094)	(0.078)	(0.079)
Obs.	224	224	224
R^2	0.005	0.357	0.357
F-statistic	1.04	50.28	35.00

Note: Robust standard errors in parentheses. All variables are standardized. Time choice represents participants' planned transcription proportion from day 1. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table A.4: *Autoplay* condition and number of videos watched

	Nb. of Videos Watched (z-score)		
	(1)	(2)	(3)
Autoplay	0.051	0.131	0.132
	(0.135)	(0.109)	(0.109)
Time Choice (z-score)		−0.590***	−0.555***
		(0.060)	(0.084)
Autoplay × Time Choice			−0.077
			(0.118)
Constant	−0.023	−0.061	−0.059
	(0.090)	(0.075)	(0.075)
Obs.	224	224	224
R^2	0.001	0.347	0.348
F-statistic	0.14	48.91	34.21

Note: Robust standard errors in parentheses. All variables are standardized. Time Choice represents participants' planned typing proportion from day 1. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.