SwitchTube: A Proof-of-Concept System Introducing "Adaptable Commitment Interfaces" as a Tool for Digital Wellbeing

Kai Lukoff
Santa Clara University
Santa Clara, CA, USA
klukoff@scu.edu

Ulrik Lyngs
University of Oxford
Oxford, UK
ulrik.lyngs@cs.ox.ac.uk

Karina Shirokova
University of Washington
Seattle, WA, USA
karina27@uw.edu

Raveena Rao
University of Washington
Seattle, WA, USA
rraveena@uw.edu

Larry Tian
University of Washington
Seattle, WA, USA
ganlin@uw.edu

Himanshu Zade
University of Washington
Seattle, WA, USA
himanz@uw.edu

Sean A. Munson
University of Washington
Seattle, WA, USA
smunson@uw.edu

Alexis Hiniker
University of Washington
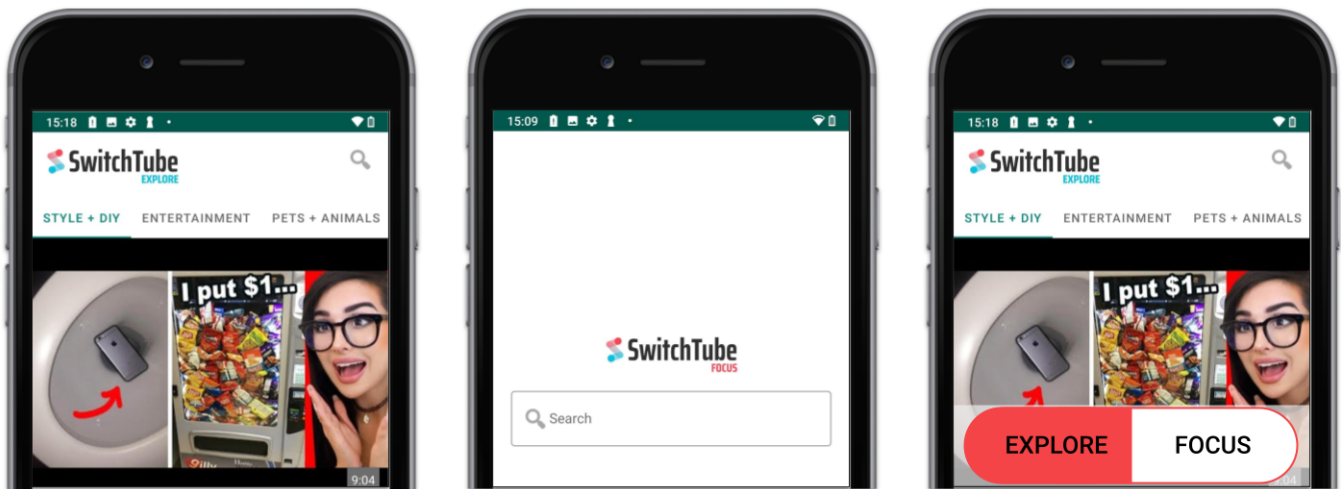Seattle, WA, USA
alexisr@uw.edu

Figure 1: The three different versions of the SwitchTube study app are designed to support different levels of sense of agency for the user. The Explore Version (low sense of agency) has a recommendations-first homepage whereas the Focus Version (medium) has a search-first homepage. The Switch Version (high) is an "adaptable commitment interface" that lets users toggle between Explore and Focus Mode.

ABSTRACT

YouTube has many features, such as homepage recommendations, that encourage users to explore its vast library of videos. However, when users visit YouTube with a specific intention, e.g., learning how to program in Python, these features to encourage exploration are often distracting. Prior work has innovated 'commitment interfaces' that restrict social media but finds that they often indiscriminately block needed content. In this paper, we describe the design, development, and evaluation of an 'adaptable commitment

interface,' the SwitchTube mobile app, in which users can toggle between two interfaces when watching YouTube videos: Focus Mode (search-first) and Explore Mode (recommendations-first). In a three-week field deployment with 46 US participants, we evaluate how the ability to switch between interfaces affects user experience, finding that it provides users with a greater sense of agency, satisfaction, and goal alignment. We conclude with design implications for how adaptable commitment interfaces can support digital wellbeing.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **HCI theory, concepts and models**; **Interactive systems and tools**.

KEYWORDS

digital wellbeing, social media, YouTube, commitment devices, adaptable interfaces

1 INTRODUCTION

Temptations, i.e., desires that conflict with long-term goals, are common in everyday life. When people attempt to resist everyday temptations, they report a failure rate of about 30% for smoking, 30% for drinking alcohol, and 35% for shopping. However, the temptation with the highest self-control failure rate of all is media use, e.g., watching TV or checking social media, which people report failing to resist 76% of the time according to the results of a day-reconstruction study [17]. Media use is again associated with a far higher failure rate than other temptations in a similar study using the experience sampling method [30].

Users often blame themselves for a lack of self-control [41], yet the alluringness of media use is no accident. Social media is intentionally designed by companies who have a financial interest in capturing and monetizing user attention [68]. As Tristan Harris, who co-founded the advocacy group Time Well Spent, notes, *"there's a thousand people on the other side of the screen whose job is to break down whatever responsibility I can maintain"* [6]. Individually the costs of these short-term distractions may seem small, but collectively they undermine people's ability to spend time in accordance with their values [65, 72].

YouTube, one of the most-watched video services in the world [42], is a prominent example of a service that is intentionally designed as a minefield of media use temptations. For example, YouTube has a rich selection of educational content, but when users go to watch it the recommender system often shows more entertaining videos that distract them away [35]. And even when users do want entertainment, the system often sucks them into

the YouTube 'rabbit hole' of watching more than they intended [66]. Specific features within YouTube—such as autoplay, related videos, and homepage recommendations—systematically induce this diminished user sense of agency, i.e., an individual's experience of being the initiator of their actions in the world [44].

In response, designers have innovated 'commitment interfaces' in which the user sets a goal of limiting time on distracting apps or services and the interface holds them to it [46]. The 'screen time tools' that now come pre-installed on all iOS and Android devices function this way. Unfortunately, this approach also indiscriminately blocks access to content and features within apps that users still want to access (e.g., educational tutorials on YouTube or the groups feature in Facebook) [43].

Instead of forcing users into an all-or-nothing bargain, one alternative is to offer an *adaptable commitment interface*, in which the user can toggle between different interfaces to meet their situational needs. However, a key question is whether such an adaptable interface provides a sufficient enforcement mechanism to hold the user to their commitment, or whether the ability to easily switch to a version with temptations will undermine the user's digital self-control [47].

In this research, we first surveyed a general population of 606 YouTube users from the US to understand their goals and motivations for changing their use. About half of participants were actively trying to change their YouTube use or planned to do so within the next month: about one-third of participant goals were about reducing quantity of use, while about half were about shifting the *quality* of videos consumed (e.g., watch fewer comedy videos, but more educational ones). Survey participants who wanted to change their YouTube use were further invited to participate in a field experiment with the SwitchTube mobile app.

We designed and developed three versions of SwitchTube: Explore (recommendations-first), Focus (search-first), and the Switch Version, an adaptable commitment interface, in which users can toggle between the Explore and Focus interfaces (Figure 1). In a three-week field experiment, 46 participants used each of the these three versions for one week. Triangulating log data, experience sampling responses, and exit interviews, we identified four key use cases in which toggling between interfaces helped users. For example, users valued being able to start with a focused search and then explore from there, rather than *immediately* landing on a homepage full of juicy recommendations unconnected to their intention. The data also revealed that the Switch Version generally provided users with the 'best of both worlds' in terms of user experience: the ability to focus on goals, *without* severely restricting their agency to explore new and satisfying content.

This paper contributes (1) the results of large survey that helps to understand how social media users want to shift their use, not just reduce it; (2) the design, development, and evaluation of a proof-of-concept system demonstrating that adaptable commitment interfaces can improve both short-term user satisfaction *and* long-term user goal alignment in the wild; and (3) implications for how

Table 1: A taxonomy of the effects of media consumption. Adapted from [25]

	Satisfaction (short-term)	Wellbeing (long-term)
Quantity of TV consumption	Time choice effects, e.g., regret for lost time	Lifestyle effects, e.g., loss of sleep
Quality of TV consumption	Content choice effects, e.g., dissatisfaction with an episode	Cultivation effects, e.g., fostering increased materialism

to design adaptable commitment interfaces to support user sense of agency.

2 BACKGROUND AND MOTIVATION

We briefly review related work on problematic YouTube use and then consider the opportunity and challenges for adaptable commitment interfaces as a new design approach.

2.1 YouTube as a Site of Problematic Technology Use

YouTube is a source of entertainment and education, but for many also of distractions that conflict with their preferences and goals [2] and lead to a sense of regret [13]. The authors of a case study of problematic YouTube use write, *"[Adam] reflects that he was initially attracted to YouTube for its educational value, but 'instead, I focused on music videos, funny videos and for entertainment purposes only"* [35, p.273]. In line with this quote, past work suggests that users have concerns over not only the quantity of content they consume, but also the quality (e.g., the ratio of entertainment to educational videos they watch) [1, 23]

In a study of television consumption [25], Gui and Stanca introduce a taxonomy for understanding quantity and quality in both the short-term and long-term as dimensions of problematic media use (Table 1). Current 'screen time' tools (e.g., Apple's Screen Time) tend to address the quadrant for time choice effects, by limiting time spent on phones or in certain apps. Yet they largely do not alter the content that users encounter within apps [44], although the aforementioned work on YouTube suggests that the quality-related quadrants are also important for digital wellbeing designers to consider. This includes both short-term 'content choice' effects such as the way that watching a video makes one feel afterwards and long-term 'cultivation effects' such as how the content of television shapes one's values and preferences over time. As a starting point, designers might benefit from a better understanding of *how* users want to change the quality of the content they consume on YouTube so that they can create tools that help users focus on the content that supports their short-term satisfaction *and* long-term well-being.

2.2 Commitment Interfaces for Digital Self-Control

Recognizing that many users struggle to manage their social media use, design practitioners and researchers have created hundreds of digital self-control tools to support them [47]. These tools leverage many different behavior change strategies, including raising

awareness of (over)use [15, 34], applying social pressure and accountability [36, 37], and affirming the user's goals and values [27, 48].

But perhaps the most common type of tool is what can be called a 'commitment interface,' in which the user commits to a goal of changing their use of a particular feature or app and the tool holds them to it [9, 56]. That is, the present self engages in reflective 'System 2' thinking and anticipates that their future self, when engaged in automatic 'System 1' thinking, will be unable to stick with their goal [31]. So the present self delegates their behavioral goals to a digital self-control tool (Figure 2). The digital self-control tool then acts as a commitment device: it enforces the wishes of the present self upon the future self. For example, a user might resolve to limit their Facebook use to 30 minutes a day and install a browser extension that locks them out when time is up.
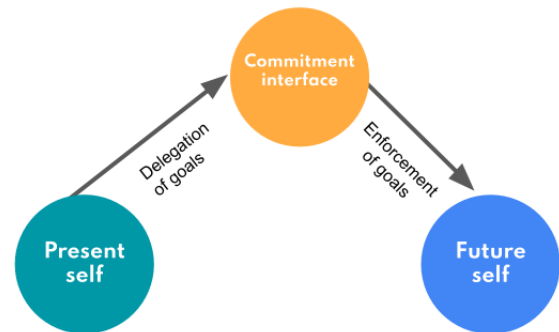

Figure 2: The present self does not trust the future self to stick to their goal, so they delegate their goal to a digital self-control tool which acts as a commitment interface. The tool enforces the goals upon the future self. But if the enforcement is too easy to override, will the commitment still be effective? Reprinted from [43]

In a review of 367 digital self-control tools, Lyngs et al. found that about 50% of tools on the Chrome browser store minimized distracting features (e.g., hiding Facebook's newsfeed) and about 60% of tools on Google Play imposed blocks on apps [47]. Among tools created as part of academic research, LocknType used a lock-out task, forcing the user to type an arbitrary string of digits in order to access distracting apps [33]. TypeOut combined a lockout task with a self-affirmation task, finding that the two interventions had an additive effect in reducing phone usage [69]. GoalKeeper imposed more significant restrictions on use, setting a time goal

and keeping users locked out of their phone at successively longer intervals when it was exceeded [32]. HabitLab tested interventions with different levels of challenge, finding that users tend to start with the difficult ones, but gradually move to easier ones, although they never give up hope that they will return to the more difficult ones [39].

A universal challenge for 'commitment interfaces' is finding the right level of strictness. Prior work has found that digital self-control tools are generally more effective at curbing screen time when the severity of enforcement or the degree of friction of the tool is stronger [16, 32], an intuition that users themselves hold [63, 69]. Too weak and it might be too easy for the user to circumvent their original goal, for example by clicking "Ignore Limit" on a warning that time is up (Figure 3a). But too strong and it might trigger frustration and lead them to abandon the tool completely, as in software that blocks a website with no override option (Figure 3b). This motivates a search for a "Goldilocks" level of enforcement that strikes a good balance: providing the right amount of goal support without leading to psychological reactance against the tool [32, 46]. We now turn to adaptable commitment interfaces as a new design approach for navigating this challenge.
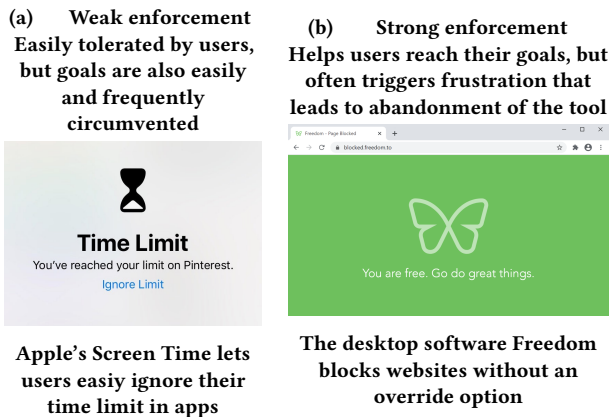
(a) **Weak enforcement**
Easily tolerated by users, but goals are also easily and frequently circumvented

(b) **Strong enforcement**
Helps users reach their goals, but often triggers frustration that leads to abandonment of the tool

Apple's Screen Time lets users easiy ignore their time limit in apps

The desktop software Freedom blocks websites without an override option

Figure 3: Enforcement mechanisms in commitment interfaces

2.3 Adaptable Commitment Interfaces for Digital Wellbeing

Adaptable interfaces are interfaces that the user can change to fit their situational needs [26]. For example, in the Microsoft Word desktop app the user can manually select "Focus Mode" to hide the toolbar and menus and write without distraction (Figure 4). As another example, the Google Maps mobile app automatically turns on 'Dark Mode' after sunset to make it easier to see driving directions at night. But the concept of 'adaptable interfaces' might also be applied beyond just productivity tools.

Many social media users also have a need for different levels of control at different times. For instance, past work has found that when YouTube users have a specific intention in mind, they tend
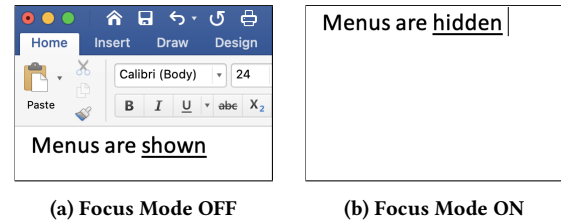

(a) Focus Mode OFF   (b) Focus Mode ON

Figure 4: The Microsoft Word desktop app enables the user to manually turn on "Focus Mode", which hides the menus and toolbar.

to prefer a search-first interface. But when users do not have a specific intention in mind, they prefer to turn control over to YouTube with a recommendations-first interface [44]. One potential solution then is to let the user change the interface to support the level of control that they want in that particular situation. This might be for a single feature, e.g., YouTube currently offers an on/off toggle for autoplay. Or it might be for a bundle of multiple features, e.g., switching between an Explore Mode and a Focus Mode.

When an adaptable interface supports the user in exercising different levels of control over temptations, we might call it an adaptable commitment interface (ACI). Our review suggests competing hypotheses for this type of interface. On the one hand, an ACI might provide the best of both worlds: an interface that lets users determine which interface best suits their needs for that situation. On the other hand, if the user can too easily switch between interfaces, an ACI might undermine the enforcement mechanism that the present self uses to hold the future self to their goals. This is a unique challenge that does not apply to most cases of adaptable interfaces (e.g., Dark Mode on Google Maps, where switching to Light Mode is not a temptation). However, in the case of digital self-control tools where there is a temporal conflict in user goals, it remains an open question whether an adaptable interface still functions as an effective commitment device.

2.4 Research Questions

We address three research questions regarding YouTube use and adaptable commitment interfaces:

- **RQ1:** *How, if at all, do U.S. YouTube users want to change the quantity and quality of their video consumption?* We draw upon the stages of change model of motivation [57] to understand how strongly and in what direction people want to change their YouTube use.
- **RQ2:** *When and why do people switch in an adaptable commitment interface?* In the Switch Version of the app, users can toggle between Focus and Explore Mode, which are designed to support different levels of sense of agency. Since users have the choice to switch at any time, will they switch in ways that undermine their intentions? Or will they switch in ways that meet their situational needs?
- **RQ3:** *How does an adaptable commitment interface influence user experience?* In particular, how do the three versions

influence measures of short-term satisfaction and long-term goal alignment?

We first report on a survey of the change motivations that YouTube users hold in the first place (RQ1), before turning to the design and deployment of SwitchTube, an adaptable commitment interface designed to answer RQ2 and RQ3.

3 SURVEY: STAGES OF CHANGE FOR YOUTUBE USERS

In behavior change, the stages of change model (also known as the transtheoretical model) is often used to assess an individual's readiness to make a change [57], ranging from pre-contemplation (no intention to change behavior) to action (actively trying to modify behaviors). We draw upon this model to answer **RQ1**: *How, if at all, do U.S. YouTube users want to change the quantity and quality of their video consumption?* We conduct a large-scale survey to assess what percentage of users are motivated to change and the nature of their digital wellbeing goals. This information can inform the design of alternative versions of social media and YouTube, as we do with the SwitchTube app in this work.

3.1 Methods

Recruitment: To obtain a general sample of users of the YouTube mobile app, we recruited participants from Prolific, a platform that specializes in crowdsourcing participants for research studies. Participants were invited to a survey about their *"YouTube watching habits and preferences."* We sought to recruit a general population of YouTube users, so intentionally did *not* advertise for participants who wanted to change or reduce their YouTube use. Participants were required to meet four inclusion criteria: live in the United States, be fluent in English, use YouTube at least once per month, and use an Android mobile phone.

Demographics: A total of 606 participants met the inclusion criteria and completed the survey (see demographics in Table 2). We excluded responses from an additional 90 participants who started but did not complete the survey. We happened to oversample young people relative to general U.S. population [64].

Table 2: Demographics of the 606 survey participants

Gender identity	Man (53%), Woman (47%), Non-binary (0%), Prefer not to say (0%)
Age range	18-24 (19%), 25-34 (39%), 35-44 (25%), 45-54 (9%), 55+ (8%)
Education	High school (34%), Associate degree (22%), Bachelor's degree (46%), Advanced degree (11%)
Race/ethnicity	White (64%), Black (12%), Asian (9%), Hispanic (6%), 2 or more races (8%), other (1%)

YouTube Use: Survey participants spent a median of 90 minutes per day (interquartile range: 40-180) on YouTube across all devices in the week prior to the survey (self-estimated[1]). Of this time, participants estimated they spent a median of 52% (interquartile range: 20-80%) in the mobile app.[2] Upon multiplying time on all devices with the percentage spent on mobile, for each participant, we found that participants spent an average of 34 minutes per day in the YouTube mobile app. This is below the average for all YouTube users: in 2017, YouTube shared that signed-in users spend an average of more than 60 minutes per day in the mobile app [49].

3.2 Procedure

In our online survey, participants first answered background questions about their demographics, technology use, and YouTube use specifically.

To investigate **RQ1**, we adapted a questionnaire from prior behavior change research [67] to assess participants' stage of change with regard to their YouTube use (Table 3). For example, our first question asked, *"Are you currently trying to take more control over how you spend time on YouTube?"* If the participant answered *"Yes"*, they were categorized into the Action Stage. If they answered *"No"* or *"Not sure"*, the survey asked further questions about their readiness to change.

To understand the nature of their intended change, we asked, *"What, if anything, do you want to change about your YouTube use?"* and how important that change was for them (5-point scale; Not at all important - Extremely important). Participants were paid $1.37 for a survey that took an estimated 8 minutes to complete, an incentive rate that exceeds the US minimum wage ($7.25 per hour). This research was approved by the University of Washington IRB.

3.3 Coding Reliability Thematic Analysis

We conducted a coding reliability thematic analysis [8] to identify commonalities in participant goals for change. Three authors read through 100 of the 606 responses together and discussed possible codes informed by Gui and Stanca's framework of the quantity and quality of media overuse [25]. They created a shared codebook with mutually inclusive codes and subcodes. For example, the change response, *"I would like to watch more educational videos rather than a lot of the junky videos that pop up on my feed"* was coded as "Increase type of content (code) → Educational or useful (subcode)" and also as "Decrease type of content (code) → Recommendations (subcode)." Two authors then applied this codebook across a random sample of 50 further responses. The interrater reliability for our codes and subcodes ranged from Cohen's Kappa of 0.66 (substantial agreement) to 1 (perfect agreement) [40]. These two authors then used the codebook to each code half of the remaining responses.

---

[1]Self-estimates of time spent on social media are only moderately correlated with actual usage [21]. We considered asking participants to report the weekly Time Watched stats presented by the YouTube app, however the help page for YouTube currently states, *"Due to a known issue, time watched on computers is reported incorrectly"* without further explanation, rendering this data source highly questionable.

[2]For comparison, the YouTube press page states that mobile accounts for over 70% of watch time [61].

Table 3: The questions and responses used to categorize survey participants into stages of change with regard to their YouTube use.

Question	Stage of change	Participants (%)
Are you currently trying to take more control over how you spend time on YouTube? Yes ———————————————————————→ No or Not sure » (go to the next question)	Action	40%
Do you plan to [...] in the **next month**? Yes ———————————————————————→ No or Not sure » (go to the next question)	Preparation	8%
Do you plan to [...] in the next **six months**? Yes ———————————————————————→ No or Not sure » (go to the next question)	Contemplation	4%
Do you **currently** feel in control of how you spend time on YouTube? Yes ———————————————————————→ No or Not sure » (go to the next question)	Maintenance Precontemplation	45% 3%

Table 4: Our coding of the changes to YouTube use that the 606 participants in our survey wanted to make.[3]

Goal Code	% of codes	Sub-code	% of parent code
Increase quantity of time	5%		
Decrease quantity of time	37%	During work	8%
		During nighttime	7%
Increase type of content	28%	Meaningful	12%
		Educational or useful	34%
		Entertaining, funny, interesting	12%
		New or diverse	18%
		Other	25%
Decrease type of content	23%	Ads	20%
		Meaningingless	27%
		Entertaining, funny, interesting	7%
		Recommendations	15%
		Other	23%
Increase awareness of use	2%		
Other	8%		

3.4 Result and Analysis

Stages of change. Based on participant survey responses (Table 3), we categorized participants into 5 different stages of change:

(1) **Action** (40%): actively trying to change
(2) **Preparation** (8%): planning to change shortly
(3) **Contemplation** (4%): wanting to change, but with no immediate plans
(4) **Precontemplation** (3%): no intention of changing
(5) **Maintenance** (45%): maintaining current behavior

Responses followed a bimodal distribution, with the vast majority of participants either at the Action Stage (40%) or the Maintenance Stage (45%). That is, participants were predominantly either actively trying to make a change or already satisfied with their behavior. We missed the opportunity to ask participants in the maintenance stage whether they had previously made a change that they were now maintaining or whether they had simply been satisfied with their YouTube use all along. Overall, about half of this sample of general YouTube users were either actively trying or in preparation

---
[3]Subcodes that accounted for less than 5% of the parent code are not displayed.

to make a change to take more control of their YouTube use.

The concept of "lagging resistance" documented for Facebook users [3] suggests that many social media users might fall into the murky middle of wanting to change but not doing so just yet. However, our results find that YouTube users who wanted to change were trying to do so now or imminently. We speculate that this may be because YouTube users face fewer *social* pressures, whereas Facebook users often feel like they cannot reduce their use because friends, family, and colleagues expect them to be active and responsive on Facebook. The fact that about half of participants were in the action or preparation stage aligns with previous findings of an underlying discontent with social media use [41].

Nature of Goals. Of the 606 survey participants, 71% shared goal(s) for change and 29% did not want to change anything. In other words, even among the 45% of participants who were in the maintenance stage and already felt in control of the time they spent on YouTube, there was still often a desire to change *something* about

their YouTube use.

Participants shared a total of 817 goals and wanted to change both the quantity and quality of time that they spent on YouTube (Table 4). In terms of quantity, participants overwhelmingly wanted to decrease the amount of time they spent on YouTube (37% of goals) rather than increase it (5%). Participants mentioned working hours and bedtime as particular times that they wanted to reduce use: *"I tend to watch it a bit too much, especially as background noise while I work from home."* In terms of quality, participants mentioned some content that they wanted to watch more of (28% of goals) and other content that they wanted to watch less of (23%). Participants wanted to watch more content that was educational (34%), new (18%), entertaining (12%), and meaningful (12%). Participants wanted to watch less content that was meaningless (27%), ads (20%), recommendations (15%), and entertaining (7%). Less common goals included increasing awareness of use (2%) and a diverse category of other goals (8%) such as creating videos, re-organizing saved videos, and reducing data usage.

Of note is that the screentime tools that come pre-installed on almost all smartphones (Apple Screen Time and Google's Digital Wellbeing) currently address only about 40% of the change goals that participants had. Given their focus on limiting time spent and presenting usage stats, they might support the goals of decreasing overall quantity of time (37% of goals) and increasing awareness (2%). However, they do not have features that support participants to change the type of content they consume within apps such as YouTube. For these goals, participants have to turn to the features of YouTube itself, where previous research has found that users feel frustration over their inability to customize the quantity, content, and placement of recommendations that appear in the interface [44].

Overall, about half of participants were actively trying to change or planning to do so shortly. A majority of these participants wanted more granular control over the type of content they consumed on YouTube, not just the ability to limit the quantity of time that current screen time tools provide. We aimed to address the needs of these participants through the design of the SwitchTube app.

Finally, the survey also served to screen and recruit participants into the next phase of our study, the field experiment. In therapy, the stages of change are often used to match the therapeutic process to the motivational stage of the client [54]. Similarly, by enrolling only survey participants seeking to make a change in our field experiment, we could ensure that our technological intervention (SwitchTube) was likely to serve the needs of our study participants.

4 SWITCHTUBE FIELD EXPERIMENT: EVALUATING AN ADAPTABLE COMMITMENT INTERFACE

Drawing upon prior work [44] led by many of the same authors of this current work and our survey results, we set out to understand how to design a mobile app that provides users with different levels of sense of agency over their YouTube experience. Two of the study

authors, together with four students in an advanced-degree technology design program, used a design process of ideating, prototyping, building, and piloting before we evaluated SwitchTube in the field.

4.1 Preparatory Design Work

Ideating. We started by ideating 10 different design dimensions to manipulate user sense of agency over time spent, drawing upon attention capture dark patterns that have been previously proposed [44, 51, 52, 70]. For example, Zagal et al. [70] introduce "playing by appointment" wherein users are required to return to game within a fixed amount of time or else lose a reward. This led us to ideate the Time Pressure dimension, which ranged on a spectrum from no control to full control. Another dimension, Content Selection, varied from maximum to minimum temptation level. We then translated each of these dimensions into 23 sets of three concrete feature ideas each that ranged along this spectrum in terms of how much support they offered for user sense of agency (some dimensions inspired multiple feature sets). For example, for the Time Pressure dimension, we imagined video recommendations that expired if not watched within 30 minutes (low sense of agency), ones that expired within a day (medium sense of agency), and ones that were always available (high sense of agency). For Content Selection, we imagined a search algorithm that was tweaked to show results with a maximum entertainment level regardless of the user's actual query (low sense of agency), one version that showed both entertaining and relevant results (medium sense of agency), and one version with only relevant results (high sense of agency). As a group, we then scored these feature sets in terms of expected impact, novelty, and technical feasibility.

Prototyping. Paper mockups for the seven highest-scoring feature sets were evaluated in 13 co-design sessions with YouTube users, as described in our prior work [44]. For example, Figure 5 shows the prototype for Content Selection with three different versions of search results. We initially anticipated building three different versions of SwitchTube along a spectrum in terms of their support for sense of agency (low, medium, high) to find a "Goldilocks" level of control as has been suggested in prior work on lockout mechanisms [32, 46]. However, co-design sessions with YouTube users revealed that rather than having a stable preference at all times, users wanted different levels of control for different situations. For example, when they had a specific intention in mind, they preferred a search-first interface, whereas when they just want to relax or pass the time, they preferred a recommendations-first interface [44]. Taking these findings into account, we instead designed two versions that support different levels of sense of agency, and a third version where users could switch between the two (hence the name SwitchTube). It also prompted us to consider that Switch (rather than Focus) might offer the highest sense of agency of all of the versions, a hypothesis that we test in this work.

We created an interactive mockup of our complete SwitchTube design in Figma and conducted usability testing with 4 participants, all university students who were active YouTube users. Participants completed four tasks, which helped us identify a number of smaller usability issues. One of the usability testing participants said they

Figure 5: For the Content Selection dimension, we showed paper prototypes of three different versions of search results that varied entertainment level to support a low, medium, or high sense of agency. Here, when a user searches for "How to cook a turkey" they see either what we called viral results (e.g., "Cute Dog Cooks Thanksgiving Dinner") or relevant results (e.g., "Easy Roast Turkey for Beginners"), and one of the versions gives the user to switch between the two types of results. Based on the user feedback we received on these prototypes, we hypothesized that the Switch Version might actually provide the greatest sense of agency.

Table 5: Three Versions of SwitchTube

Version	Intended level of sense of agency	Homepage	Search Results	Video Player
Explore	Low	Unlimited recommendations (from 8 US YouTube categories, e.g., music, comedy)	Viral results - append "viral" to the user's search query	• Next video autoplays • Related videos appear underneath
Focus	Medium	Recommendations off by default (user can individually choose categories to turn on)	Relevant results - standard YouTube results for the user's query	• No autoplay • No related videos
Switch	High	Toggle lets the user switch between Focus Mode and Explore Mode.		

would like to use the low-agency version to *"explore viral content"* and the high-agency version to *"focus on my goal,"* which led us to call the two versions "Explore" and "Focus." While labeling the two versions in this way could lead study participants to form preconceived notions of how to use that version (as opposed to say,

"Version A" and "Version B"), we decided that this was worthwhile because it would make it easy for participants to recall the two versions in the exit survey and interview and seemed unlikely to

signal to participants that they *should* necessarily prefer one version over the other.

Table 5 shows an overview of the three different versions of the final SwitchTube study app and their features. A screenshot of the homepage of each of the three versions is shown in Figure 1. A short video introduction and captioned screenshots of the entire app are also available on the Open Science Framework: https://osf.io/z735n. We refer to the three versions of the app as the Explore Version, Focus Version, and Switch Version, and the two toggle options *within* the Switch Version as Explore Mode and Focus Mode.

Our aim for Focus was to support the user's specific intention for visiting the app if they had one (e.g., learning how to cook a turkey), whereas Explore was designed to maximize distractions that would take them away from their original intention. In doing so, we expected that sense of agency (the user's experience of being the initiator of their actions) would be supported in Focus and diminished in Explore. To this end, we appended "viral" to every search query submitted in the Explore Version. Our goal was not to fill Explore with viral content per se, but rather to add noise *and* temptations to the user's search results. To simply add noise, we could have appended *any* term to the user's search query (e.g., "zebras"), but our internal testing of several different terms (e.g., "entertaining," "funny," and "creative") suggested that "viral" was the most effective at returning results that were also tempting.

Homepage video recommendations in SwitchTube were not personalized due to restrictions of the YouTube Data API (personal watch history and recommendations are difficult-to-access due to understandable privacy concerns). We further address the absence of personalized recommendations in the discussion section. Instead, homepage recommendations were drawn from the most popular videos in different YouTube categories (e.g., music, comedy) for the U.S. region. These are the same non-personalized recommendations that are displayed in YouTube's own categories. In the Explore Version, the homepage featured an unlimited scroll of these videos and the video player showed related video recommendations below the video that was currently playing and autoplayed the next related video. In the Focus Version, the homepage hid recommendations by default and related videos and autoplay were removed by design.

Building. An illustrated software architecture model for the SwitchTube study app on Android is shown in Figure 6. The app assigned participants to experimental conditions and used a logger to monitor information about how participants used the app. The user interface had homepage video feeds, a video player, and search results. These were populated with data pulled from the YouTube Data API and the Google Custom Search API. Finally, the app conducted experience sampling, which we built as a custom system. All of this data was sent to the Firebase Realtime Database and then synced with Google BigQuery to allow for custom views and further analysis.

One particular challenge we encountered was a severely limited quota for the YouTube Data API, which we needed to populate the video recommendations on the homepage and video player (related videos and autoplay) and return search results. When we built SwitchTube, YouTube restricted developers to a default quota

of 10,000 per day, whereas default quota at the time of previous research had been 1 million at the time of previous YouTube research [28]. As a result, we quickly maxed out our quota in our testing of the app (e.g., a single search has a quota cost of 100). We tried the official form for requesting an increased quota, but received no response. Drawing upon our privileged position, we contacted multiple personal connections at Google in managerial positions who were also unable to get the YouTube Data API team to grant our request. In the end, we were forced to integrate a second API into SwitchTube (the Google Custom Search API), which we could pay for and use to populate search results, but it cost us considerable time and effort to do so. Two years later and after we had finished our deployment study, we received an email that YouTube has finally launched an official YouTube Researcher Program with expanded access to their Data API[4]. We hope our report of this barrier lends support to the regulatory push to require large technology companies to provide researchers with greater access to audit and redesign their algorithmic systems for digital wellbeing.

Piloting. Our research team internally piloted the SwitchTube study app on a variety of Android devices over eight weeks. This again identified countless usability issues, from the font size of the experience sampling prompts to missing log data, that we resolved in the next version. We then recruited four students, all active YouTube users, from outside the research team for external piloting, which identified still further issues about study procedures, but also confirmed that the app was ready for deployment. We note that these participants identified several usability issues that we simply decided not to fix (e.g., when the phone was rotated horizontally the video had to reload). Our goal was not to rebuild a user experience as seamless as YouTube itself (which would have required a Herculean effort), but rather to develop a proof-of-concept system that would be acceptable enough for participants that they would engage with it sufficiently to address our research questions [22].

4.2 Pre-Registered Hypotheses

Our third research question asks how adaptable commitment interfaces influence user experience. In line with this question, we posed several specific hypotheses. Following the best practices of the open science movement, we pre-registered these *before* examining the data: https://osf.io/sevfd. This helped us think through our study protocols in advance and guard against the the natural temptation of hypothesizing after the results are known (HARKing) [14]. As noted in the pre-registration, in addition to this *confirmatory analysis* with pre-registered hypotheses we also planned to conduct *exploratory analyses* of the log data from the app, such as time spent in the different versions, but to use only descriptive statistics for this purpose.

In general, our pre-registered hypotheses tested whether or not the Switch Version (an adaptable commitment interface) provides the 'best of both worlds' across measures for sense of agency, satisfaction, and personal goal alignment. All of our hypotheses were tested based on measuring the mean per participant rating (1-7) of
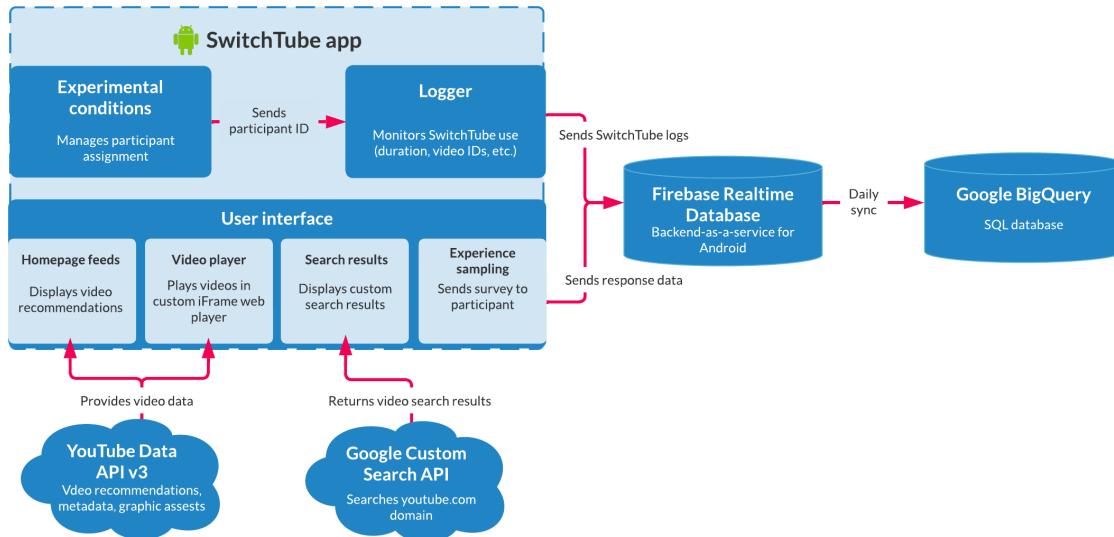
---

[4]https://research.youtube/

Figure 6: The technical implementation of the SwitchTube study app for Android.

experience sampling mechanism (ESM) responses for these metrics.

H1: User Sense of Agency. Our first set of hypotheses (H1a-H1c) addressed user **sense of agency**, which prior work suggests is at the center of user concerns with social media [4]. Our expectation was: Switch > Focus > Explore, which corresponds to 3 pairwise comparisons:

- H1a: The mean rating will be higher for Focus than Explore.
- H1b: The mean rating will be higher for Switch than Explore.
- H1c: The mean rating will be higher for Switch than Focus.

The features in Focus and Explore were based on our prior research into how the features of YouTube affect user sense of agency [44]. As Switch lets users toggle between the Focus and Explore interface, we expected that this additional option would further increase user sense of agency.

H2: Satisfaction. Our second set of hypotheses (H2a-H2c) addressed user satisfaction, as in the short-term pleasure that users derive from social media apps. Our expectation was: Switch > Explore > Focus, again corresponding to 3 pairwise hypotheses that follow the same pattern as H1. In our previous study of YouTube [44], users reported that homepage recommendations often provided short-term satisfaction, but Focus hides these by default. We expected Switch might provide a useful option to avoid recommendations at times when they are not wanted.

H3: Goal Alignment. Our third set of hypotheses (H3a-H3c) addressed personal goal alignment, as in how well app use aligned with the user's long-term goals for use. Our expectation was: Switch > Focus > Explore, which again implies 3 pairwise comparisons.

This is because our previous work found that that search often supports YouTube users' personal goals [44], but Explore minimizes the search option and adds distracting temptations to the results. On occasions where recommendations might actually better support the user than search (e.g., as in when survey participants said their goal was to find new or diverse content to watch), Switch would also provide that option.

4.3 Methods

Recruitment. We screened the 606 participants from our survey for the following three inclusion criteria:

(1) Action or preparation stage of change with regards to their YouTube use (48% of survey participants met this criterion).

(2) Own an Android smartphone with operating system version 6.x - Marshmallow or higher. This was because the study app did not support older versions (87% of survey participants met this criterion).

(3) Spend a minimum of 10 or more minutes per day on the YouTube mobile app, according to self-estimate (75% of survey participants met this criterion). This was to ensure that participants already had a regular habit of watching videos on mobile, making it more natural for them to use Switch-Tube.

This left us with 146 survey participants who were eligible to also become experiment participants. Given that a prospective power analysis for ESM studies requires an estimate of effect size that is difficult to obtain for a novel technology, we instead followed Berkel et al.'s guidance and informed our target number of participants using local standards in the HCI community [5], where the median is 18 participants and the mean is 53 [10]. Since we wanted to be

able to detect differences between conditions with a high degree of confidence using frequentist hypothesis testing, we set a target of having 45 participants complete the field experiment.

We invited eligible survey participants to participate in small batches until we approached our target. In the invitation to the study and again upon installing the study app, participants were informed that the research team would monitor and analyze their activity in the study app, including their searches and the titles of the videos they watched.

Demographics. A total of 46 participants completed the experiment (see demographics in Table 6). We happened to oversample Asian, Black, and young people relative to the general U.S. population [64].

Table 6: Demographics of the 46 field experiment participants

Gender identity	Man (54%), Woman (46%), Non-binary (0%), Prefer not to say (0%)
Age range	18-24 (35%), 25-34 (39%), 35-44 (17%), 45-54 (4%), 55+(4%)
Education	High school (35%), Associate degree (13%), Bachelor's degree (37%), Advanced degree (15%)
Race/ethnicity	White (43%), Asian (26%), Black (20%), Hispanic (7%), 2 or more races (4%)

YouTube Use. Field experiment participants spent a median of 140 minutes per day (interquartile range: 120-240) on YouTube across all devices in the week prior to the survey (self-estimated). Of this time, participants estimated they spent a median of 63% (interquartile range: 40-84%) in the mobile app. We again multiplied time on all devices with the percentage spent on mobile, for each participant, to find that field experiment participants spent a median of 87 minutes per day in the YouTube mobile app. This is considerably higher than the median of 34 minutes per day spent in the app by all survey participants, indicating that those who were invited and participated in the field experiment were heavier YouTube users.

Procedures. As shown in Table 7, participants completed an entrance survey, one week of use of each of the three versions of the SwitchTube app (Explore, Focus, Switch), an exit survey, and, for a subset of participants, an exit interview. In the entrance survey, participants completed additional questions about the nature of their YouTube use and received instructions for installing the SwitchTube study app on their Android phone from the Google Play store.

Upon installing SwitchTube, participants were assigned to start in either Explore or Focus following a counterbalanced assignment. Although it risked introducing ordering effects, we decided against also counterbalancing the Switch condition. Instead, Switch always came last so that we could understand when and why participants choose to toggle between Explore and Focus (RQ2) after having experienced each for a week. In each week, participants were required to use the app for 3 or more days for a total of at least 30

minutes. If participants did not meet these requirements, they were disqualified from further participation, but still compensated for their participation to that point.

The SwitchTube app collected both objective and subjective data. Objective data included logs of time spent, searches made, videos watched, and the source of watched videos (e.g., homepage recommendations). In terms of subjective data, participants were experience sampled using the three questions in Table 8. Conceptually, we wanted to capture an understanding of how the different versions influenced sense of agency, as well as satisfaction in the sense of short-term pleasure and goal alignment in the sense of long-term personal goals. Unfortunately we could not find validated scales that were short enough to be suitable for ESM, but we tested our wording for clarity in our piloting. This led us to clarify that we wanted participants to answer about this particular session of use *("For this SwitchTube use")* rather than for their use of SwitchTube as a whole.

In terms of timing, a prompt appeared with these three questions on the participant's phone when the following conditions were met:
(1) The participant had not already responded in the past hour;
(2) The participant had used the app for at least 30 seconds;
(3) The app went into the background (e.g., the user exited the app to the phone's home screen or they switched to another app).

If the participant did not respond within one minute, the prompt disappeared.

After completing one week each in Explore, Focus, and Switch, participants completed an exit survey. In the exit survey, participants were shown screenshots of the homepage, search results page, and video player in Explore and Focus as a reminder and answered which they preferred and why. Participants then explained when and why they switched between versions of the app. Finally, they answered which version of the app they preferred and why.

Exit interviews were conducted remotely over Zoom with a subset of participants using a method called data-driven retrospective interviewing [60]. Using screen share, participants were shown counts, tables, and visualizations from their own log data, e.g., time spent in the app, occasions when they switched between versions, and their ESM ratings, and asked questions intended to elicit the "why" behind their behaviors. For example, we asked:

> In the Switch version, you switched between the Focus and the Explore interface 17 times. Can you look at the table below, choose a couple of examples, and describe why you switched at that time?

We also retrieved the original change that participants wanted to make to their YouTube use from the survey and asked them whether the different versions of SwitchTube supported that goal. A total of 16 participants were interviewed, at which point we believe we reached data saturation with regards to our research questions. Interviews lasted about 45 minutes each.

Table 7: SwitchTube Study Timeline

	Start	**Week 1**	**Week 2**	**Week 3**	**Finish**	**Subset of participants**
Study App	Install the app	Explore or Focus	Explore or Focus	Switch	Uninstall the app	
Activities	Entrance Survey	Each week, use the app on 3+ days for a total of at least 30 minutes. Answer the experience sampling questions.			Exit Survey	Exit interview
Incentive	$5	$15	$30	$50 (for Week 3 and Exit Survey)		$20

Table 8: Experience Sampling Questions

	Question	Scale
Sense of agency	For this SwitchTube use, how much did you feel out of or in control?	1=very out of control 7=very in control
Satisfaction	For this SwitchTube use, how much did you feel dissatisfied or satisfied?	1=very dissatisfied 7=very satisfied
Goal alignment	For this SwitchTube use, how much did it conflict with or support your personal goals?	1=very in conflict 7=very supported

Participant incentives were backloaded to encourage participants to complete the entire study, allowing us to compare their experience between conditions. This meant: $5 for the entrance survey, $15 for week 1 of app use, $30 for week 2, $50 for week 3 and the exit survey, and $20 for the exit interview. To protect data privacy, we assigned each participant a unique identifier (e.g., 446565) that was associated with their usage data. We connected this data to the participant's personally identifiable information (e.g., contact information) only for the exit interviews, where we presented participants with a personalized summary of their usage. This research was approved by the University of Washington IRB.

4.4 Data Analysis

Log data were analyzed in an exploratory fashion and are presented as descriptive statistics. Experience sampling data were analyzed according to our pre-registered hypotheses. Exit survey and exit interview data were analyzed together by three of the authors, who conducted a codebook thematic analysis that addressed our research questions [7]. We first read through all of the data and added initial codes using Delve, a tool for collaborative qualitative coding. Initial codes were refined and consolidated through group discussion, leading to a final codebook which was applied to the data. For example, one code was "within-session switching," which was for cases in the Switch Version where participants described a use case of toggling between Explore Mode and Focus Mode during a single visit, for example:

> I actually would switch between the two regularly, but specifically to use focus as a search bar and explore as a way to find new content I was not thinking of. (P41)

Each code and its supporting quotes was then moved to Miro, a tool that that we used to conduct affinity diagramming, in which codes were clustered according to our research questions. Through

this process, we found that there were four distinct use cases for switching. We then drew upon these codes and their associated quotes to write four analytical memos: switching behavior (for RQ2) and sense of agency, satisfaction, and goal alignment (for RQ3), which form the basis of our results.

5 RESULTS AND ANALYSIS

Our log data, experience sampling, and interview data allowed us to triangulate the experience of 46 participants in our study using both objective and subjective measures.

5.1 App Usage Data

In terms of usage of SwitchTube, there were notable differences and non-differences in how participants engaged with the three versions of the app for one week each (Figure 7). Our analysis of usage focuses primarily on the comparison between Explore and Focus, before turning to when and why participants used the toggle in Switch (RQ2) and how each version influenced user experience (RQ3).

Time spent. The median time spent using each version of the app for a week was as follows:

Time spent in the app per week (minutes)		
Explore	Focus	Switch
117	145	107

On the one hand, participants spent considerable time with our study app, far beyond the 30 minutes per week that we asked of participants. At most, one participant even spent 752 minutes (12.5 hours) in a week in the app in Switch. On the other hand, given the high usage of YouTube that our participants reported (a median of 87 minutes per day on mobile), SwitchTube still likely represented only a fraction of their overall time spent watching YouTube videos.

Surprisingly, the time spent in Focus actually exceeded the time spent in Explore and Switch, whereas we had anticipated that the opposite would be the case because of the lack of recommendations in Focus.

Time spent per video watched. The median time spent per video that participants started to watch (regardless of whether or not they finished it) was as follows:

Time spent per video watched (minutes)		
Explore	Focus	Switch
6.6	10.6	9.1

Participants spent more time on each video that they started in Focus, which partially explains why the overall time spent was greater in that version of the app.

Videos watched. The median number of videos that participants watched over the course of a week with each version was as follows:

Videos watched (#)		
Explore	Focus	Switch
16.5	11	11

Participants started to watch more videos in Explore than in the other versions, even though they spent less overall time in Explore. In other words, participants in Explore started more videos that they didn't finish as their selections were driven by recommendations rather than content than they had specifically searched for. The videos that participants selected in Explore were also slightly shorter (median 8.8 minutes) than in Focus (10 minutes) and Switch (11.9 minutes).

Searches made. The median number of searches that participants made was as follows:

Searches made (#)		
Explore	Focus	Switch
5.5	10	4.5

Participants made more searches in Focus, suggesting that (a) the lack of recommendations led to more searching and/or (b) participants preferred the relevant results in Focus over the viral results in Explore.

Time spent browsing. The median percentage of time spent browsing in the app (any time spent in the app when no video was playing), was as follows:

Time spent browsing (%)		
Explore	Focus	Switch
49%	46%	38%

Participants in Explore spent 3 percentage points more of their time in SwitchTube browsing than they did in Focus and 11 percentage points more than they did in Switch. This is likely because there were more recommendations to browse in Explore.

Videos watched by source. Participants could watch videos from the homepage recommendations, from related videos (which appeared underneath the video that was currently playing), or from searches (Figure 8). From homepage recommendations, participants watched far more in Explore (median: 6 videos) than in Focus (median: 1), indicating the strong influence of an interface where the user has to opt-in to recommendations. By contrast, Explore and the current interface of the YouTube mobile app show recommendations without even providing an opt-out choice. From related videos, participants watched a median of 3 videos in Explore, whereas these were completely unavailable in Focus. From search, participants watched far fewer videos in Explore (median: 2) than in Focus (median: 7.5). Across these three video sources, Switch was in the middle between Explore and Focus suggesting that participants used features from both versions of the app, rather than just keeping the toggle on one mode.

Videos watched by kind. In Focus, participants tended to consume videos that all stemmed from the same search and were thus closely related: for instance, P21 played "15-minute morning power yoga" [59] followed by "15-minute hands free morning yoga" [58]. By contrast, in Explore, the same participant jumped from an ASMR-relaxation video to a celebrity cooking challenge to a popular hip-hop music video, consuming a great diversity of content. In Explore, viewing was driven in large part by homepage recommendations of the most popular videos in the U.S., so videos tended to appeal to a broad audience (e.g., "If You Can Carry $1,000,000 You Keep It!" [53]) and often included creators who are celebrities on YouTube (e.g., MrBeast) and beyond (e.g., Ed Sheeran). By comparison, participants in Focus often viewed more niche videos, such as one on opening moves in chess that is unlikely to be recommended on the homepage as one of the most popular videos in the U.S.

We make available an open dataset of 2063 video play events, including the title of the video, the version of the app (Focus, Explore, or Switch), and the source of the video (search, related videos, or homepage recommendations) via the Open Science Framework: https://osf.io/z735n. To protect the privacy of study participants, no participant ID is included. While outside of the scope of this paper, future work could analyze this data to identify further differences in the nature of the content consumed in the different versions of SwitchTube that might inform the design of content algorithms for digital wellbeing.

5.2 The SwitchTube User Experience

We triangulated data from multiple sources to understand how participants experienced SwitchTube. Log data served as an objective measure of feature use, while experience sampling, the exit survey and interviews revealed contextual information and the subjective experience behind the usage.
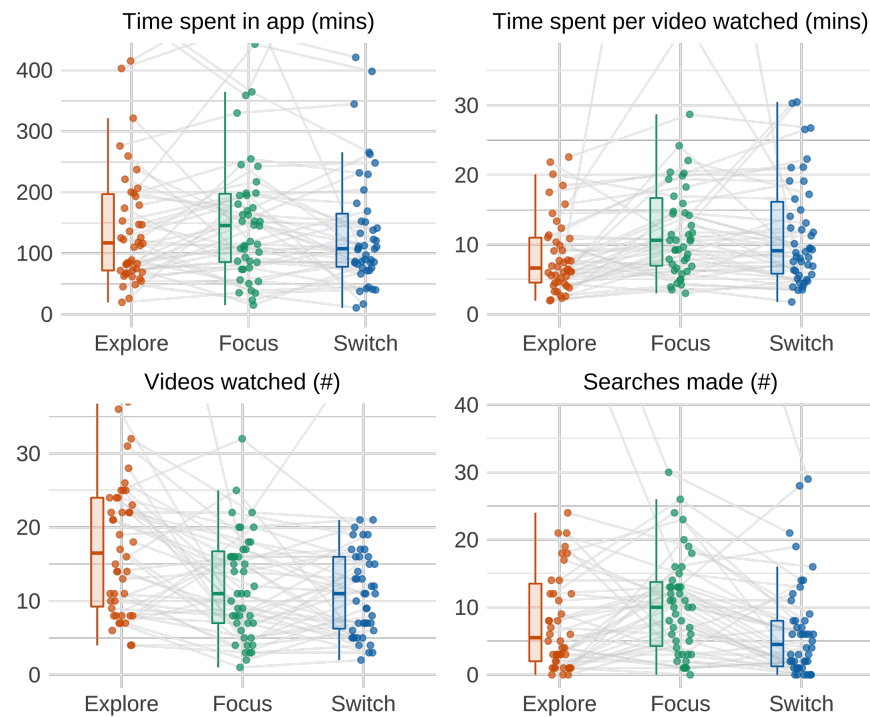
Figure 7: Log data from our deployment study with 46 participants for the three versions of the SwitchTube app: Explore, Focus, and Switch. Participants spent more time in Focus than Explore. Participants also spent more time on each video that they watched (or started) in Focus, although they actually watched more videos in Explore. Participants in Focus also made more searches than in the other versions. For a detailed explanation of the format used for this figure, which is repeated in several subsequent figures, see the footnote. [5]
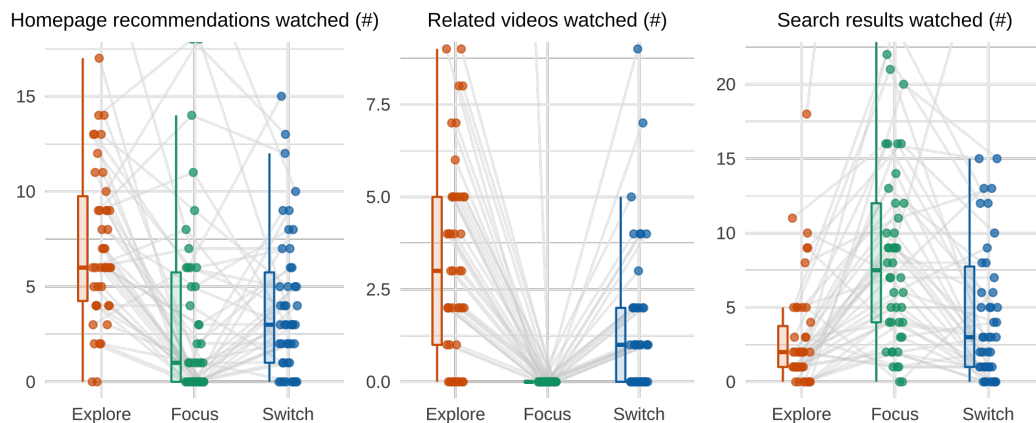

Figure 8: Videos watched from three different sources: homepage recommendations, related videos (which appeared underneath the video that was currently playing), and from searches. Video watching was driven by recommendations in Explore and by searches in Focus. In Switch, there was a balance between the two.

---

[5]The boxplot shows the median, the interquartile range (Q1-Q3) and, with the top whisker, Q3 + 1.5x the interquartile range. Each dot represents one of the 46 study participants and the gray lines show how that participant's data changed between versions of the app. To zoom in on the vast majority of the data, we calculated the 95th percentile for all three versions, took the maximum value, and cropped the top of each figure at that point; the gray lines going off the top indicate outliers that extend

above the top of the figure. The Explore and Focus conditions were counterbalanced in the actual experiment, but are always displayed Explore first, Focus second in this figure. Switch always came last, as discussed in the paper.

5.2.1 Why and When People Switched (RQ2). In Switch, users could toggle between Explore and Focus Mode. Almost all users made use

of this feature, making a median of 8.5 switches during their week in the app (interquartile range 4-13). Our analysis found that **four distinct use cases** captured switching behavior and motivation:

Use Case 1 — Curiosity about feature differences. When they first opened Switch, participants switched between the Explore and Focus Modes in rapid succession. Although they had already experienced each of these individually for a week, they were still curious about the differences between the two: *"I went back and forth a little bit just to see if anything was added to either one of them, or which one offered more" (P7).* This curiosity-driven switching not only satisfied an itch to understand the differences, but also facilitated preference discovery, i.e., helping participants decide which version was a better fit for them.

Use Case 2 — Curiosity about new content. Participants sometimes switched from Focus to Explore Mode to see recommendations:

> I would occasionally switch to Explore to see if there was anything popular—it was how I discovered The Weeknd released a new song! But for the most part I preferred to stay in Focus. (P1)

This use case aligns with one of the top goals that survey participants had: watching more new or diverse content. Sometimes recommendations inspired participants to search in a new direction: *"While I do like searching for content myself, sometimes I do get bored and looking at genres reminds me of things I would also like to search for" (P36).* In these cases, switching from Focus to Explore Mode satisfied a need for novelty.

Use Case 3 — Between-session switching: addressing situational needs. Another common use case was when a participant had a particular intention for their session of use and toggled to the most suitable interface: *"If I'm searching for something like 'flat tire' I'd go with Focus because that seems to be more educational. Explore is definitely more entertainment" (P38).* Similarly, another participant said,

> With Explore showing the most watched content, I often found what I wanted easier with it. If I was however watching something for educational purposes I always used Focus first, finding myself less prone to wasting time watching senseless videos. (P12)

In this use case, the adaptable commitment interface of Switch was effective at holding the user to their commitment.

For another participant, Focus also made them realize how strongly design patterns influenced their behavior:

> Focus was better for restricting my viewing and keeping with my restrictive schedule. In Focus, the app would let me select a video and after it was done it would end with no autoplay directly afterwards. What I learned was that autoplay was a big reason for continual watching; far more than I ever knew. (P10)

Other participants appreciated how Focus limited *"distractions,"* *"unintended videos,"* and *"losing track of time."* This pattern of use accorded with our prior research: users come to YouTube with many different intentions, so recommendations are sometimes helpful.

However, the current interface of the YouTube app aggressively promotes new content discovery, even when that is not the user's intention [44].

Use Case 4 — Within-session switching: start focused, explore from there. A final and unanticipated use case was that participants frequently toggled between Focus and Explore within the same session, as shown in Figure 9.

> I used Focus to search for a particular video and when I was done watching I would turn on Explore sometimes, when I felt like looking for recommended results underneath. (P13)

In this use case, participants started in Focus Mode to avoid distraction from homepage recommendations. Then they searched for and clicked on a relevant result. Once the video started or finished playing, they then toggled over to Explore Mode to reveal related videos. In this way, they could feed the system with their initial intention without getting distracted, and then explore from there. With an understanding of app usage data and switching behaviors, we now turn to how SwitchTube influenced user experience measures.

5.2.2 **H1: Sense of Agency.** In order to analyze the experience sampling ratings, starting with sense of agency, we first had to select an appropriate statistical method. Since our data were derived from an ordinal scale, determining whether or not the means of each participant ratings are normally distributed was important. We performed a Shapiro-Wilk test that showed that the distribution of participant means in the Switch condition departed significantly from normality (W = 0.96, p-value < 0.01). Based on this outcome, we use non-parametric statistics in all subsequent analyses of ESM ratings.

For sense of agency, our hypothesis was: Switch > Focus > Explore. The median of the participant means was 5.4 for Switch, 5 for Focus, and 4.1 for Explore (Figure 10). The participant means for control were statistically significantly different across conditions using the Friedman test of differences among repeated measures, $\chi^2(2) = 15.3$, $p < 0.001$. A pairwise Wilcoxon signed rank test with a Holm-Bonferroni adjustment for multiple comparisons revealed statistically significant differences in participant mean score between Explore and Switch ($Z = 4.04$, $p < 0.001$) and Explore and Focus ($Z = 2.71$, $p = 0.14$), but not between Focus and Switch ($Z = 1.9$, $0 = 0.059$). All subsequent ESM analyses use these same tests.

Because we had based the design of the SwitchTube app on features that participants reported as having an influence on their sense of agency, these hypotheses served as a manipulation check that was largely confirmed. The features in Focus did indeed increase sense of agency and the features in Explore decreased it.

The inability to customize recommendations led participants to feel like their attention had been led astray. One participant described not being able to find the Spanish-language music videos he was looking for in Explore:

STEP 1
Search results for "Cake decorating" in Focus Mode

STEP 2
Watch video on cake decorating

STEP 3
Related videos appear below after toggling to Explore Mode

STEP 4
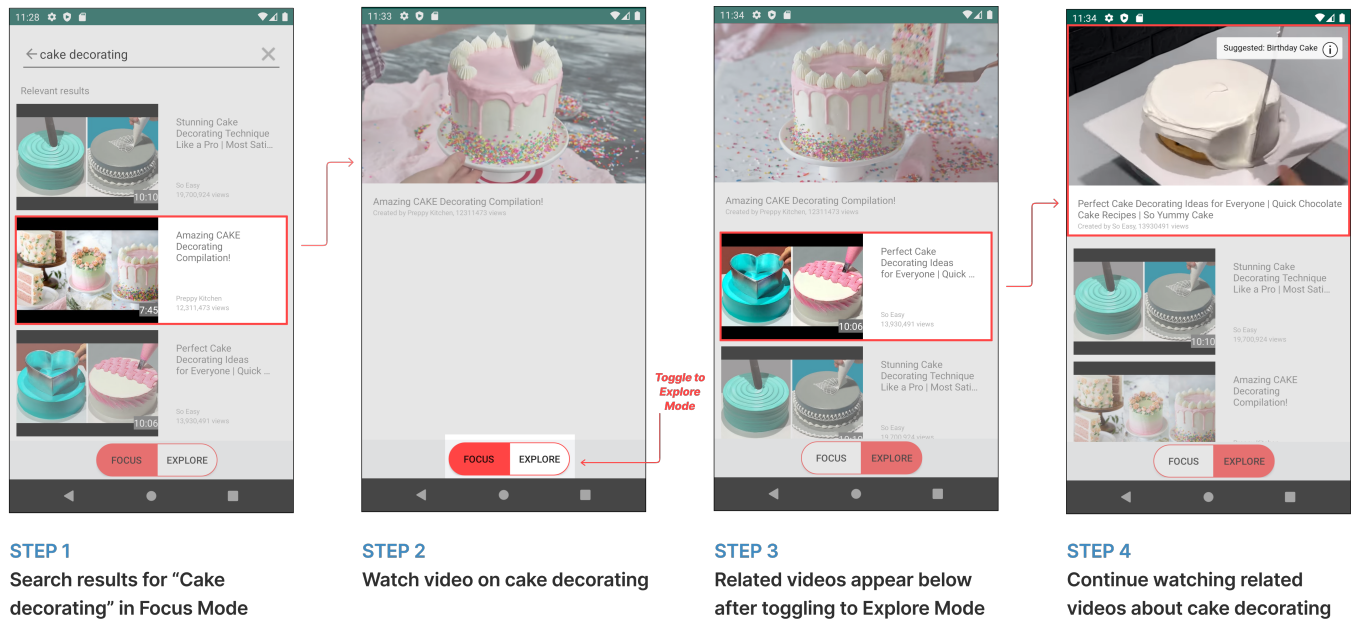Continue watching related videos about cake decorating

Figure 9: In this example of Use Case 4, recreated based on one of our exit interviews, P3 started with a search in Focus and then toggled to Explore to see related videos. By switching within a single session of use like this, participants were able to perform a search in line with their initial intention without getting distracted, but then still explore related content from there.
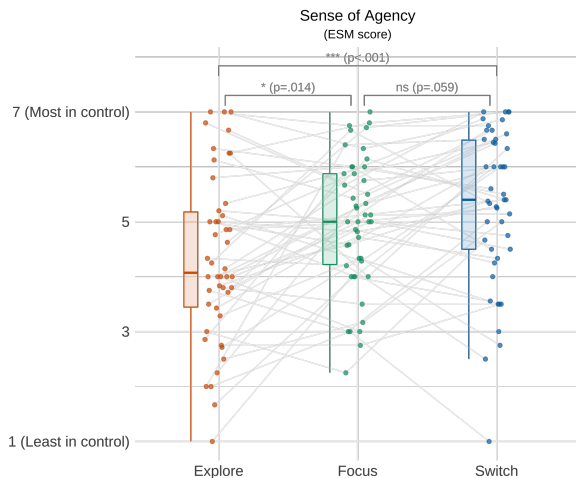

Figure 10: Experience sampling ratings for sense of agency were higher for Switch and Focus than Explore. Explore also had the greatest variation, with more participants feeling out of control as compared to the other versions. Note: Each dot represents the average of all of a single participant's ESM responses.

> I felt like I had no control, because it was only the videos that were on there and the search was not working very well at all. So I was limited to what somebody else chose for me. (P7)

This sentiment also aligns with the findings of our previous work, where participants also found the current customization options in the YouTube mobile app (e.g., marking "Not Interested" on recommendations) to be wholly inadequate in terms of providing the level of control they desired [44].

Switch was rated higher than Focus for sense of agency, but not quite significantly so ($p < 0.059$). Within Switch, the Focus Mode was generally preferred, but there were times when participants appreciated the option of toggling to Explore. In line with **Use Case 3** (between-session switching to address situational needs), one participant described:

> Oddly enough, my family collects different versions of the Monopoly board game just for fun... Since I was looking for videos [of new versions] late at night with my family, I just needed something really easy to do. Trying to fight through the search bar at that point was just too much effort. Just let me scroll through and find something interesting and we'll call it good. (P27)

In other words, on occasion participants would 'satisfice' by accepting recommendations that were suboptimal but low effort, rather than "fighting" the interface for the perfect video. While having the option of toggling on recommendations sometimes supported sense of agency, an inability to turn them off at all undermined sense of agency (as is the case in the Explore version and the current YouTube mobile app).

5.2.3 **H2: Satisfaction**. For satisfaction, our hypothesis was: Switch > Explore > Focus. The median of mean scores per participant

was 5.3 for Switch, 4.8 for Focus, and 4.2 for Explore (Figure 11). Participant means for satisfaction were statistically significantly different across conditions: $\chi^2(2) = 14.78, p < .001$. There were statistically significant differences in participant mean score between Explore and Switch ($Z = 3.63, p < .001$), Explore and Focus ($Z = 2.66, p = .016$), and Focus and Switch ($Z = 2.1, p = .037$).

Switch again rated highest, indicating that its mix of features supported short-term user satisfaction. However, contrary to our expectations, Focus was actually rated higher than Explore. Whereas we had expected that recommendations would contribute to short-term satisfaction, participants reported otherwise.
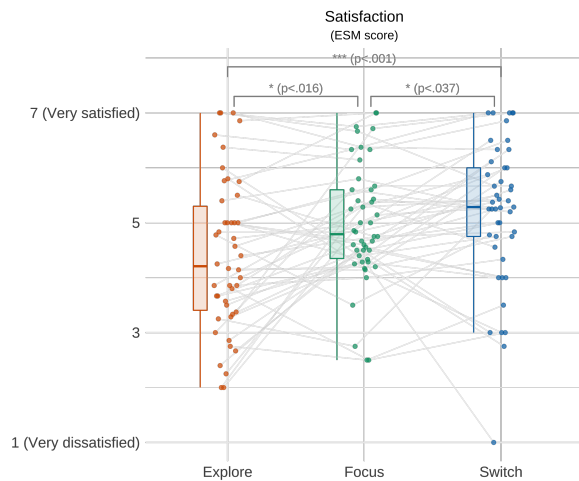

Figure 11: Experience sampling ratings for satisfaction were higher for Switch than for Focus and higher for Focus than for Explore. Again, Explore also had the greatest variation.

In interviews, participants appreciated the ability to switch between the interfaces to meet different needs. In line with **Use Case 2**, this satisfied curiosity to check out different content across the two versions:

> I switched back to Focus Mode, searched for Vivaldi, found "The Four Seasons," and played that. And then, I guess, four hours later I opened it back up and was like, 'Let's do something else,' switched to Explore Mode and watched two random videos on vending machines. (P3)

One might guess that the classical music listener is unlikely to watch videos about vending machines, but our participants did clearly enjoy exploring at times. Similarly, another user appreciated different search results for different occasions: *"Both options are cool because one is relevant to the search terms and the other one is viral videos so I think both are effective"* (P9). The option to switch provided greater satisfaction.

Linking satisfaction back to sense of agency, many participants liked having the options that Switch afforded, even if they did not exercise them: *"Switch did satisfy my needs more just because I had the option to switch between the different modes... Even if you*

don't use it, having the option is always good" (P23). In other words, just having the feeling of being in control felt good. In cases like these, our research team found it difficult to disentangle whether a quote best described a participant's sense of agency, satisfaction, or goal alignment. They were all related, which is also seen in the experience sampling ratings which follow a consistent preference order across all three of our measures: Switch >= Focus > Explore.

5.2.4 ***H3: Goal Alignment***. For H3, our expectation was: Switch > Focus > Explore. The median of mean scores per participant was 5.3 for Switch, 4.8 for Focus, and 4.3 for Explore (Figure 12). Participant means for goals were statistically significantly different across conditions: $\chi^2(2) = 13.32, p < .001$. There were statistically significant differences in participant mean score between Explore and Switch ($Z = 3.96, p < .001$) and Explore and Focus ($Z = 2.74, p = .012$), but not Focus and Switch ($Z = 1.73, p = .085$).
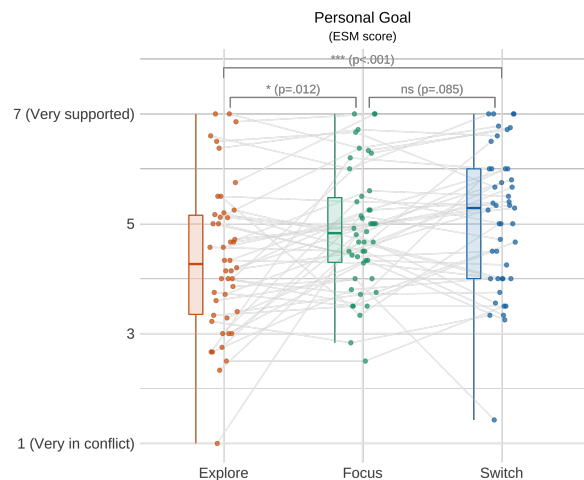

Figure 12: Experience sampling ratings for personal goal alignment were higher for Switch and Focus than Explore.

The limited number of features in Focus supported goal alignment. With just a search bar and video feed feature on the home page, the minimal design made users feel *"less likely to get distracted"* by recommendations and related videos. Instead, users felt it was easier to focus on the videos they intended to watch and spend less time browsing the app. For example, in the initial survey, P33 expressed their high-level goal as the following: *"I want to be able to use this app less than I am used to."* When interviewed, they explained that Focus best aligned with this goal:

> I liked how in Focus it was just the video and nothing else. There was no comments to read, no suggested videos on the bottom to scroll through. So, I feel like I was very focused on that one video that I was watching.

Without the distraction of comments or video recommendations that they felt *"forced"* to watch, users were able to quickly accomplish what they were set to do in SwitchTube, close the app, and do something else. Knowing that Focus does not show recommendations, participants used the app when they had a specific goal

in mind as described by P6 who wanted to reduce the quantity of time they spent browsing: *"It was more, set a goal, find it, and complete that goal instead of just scrolling through randomly looking for something."*

In Switch, participants commonly described experiences that aligned with **Use Case 4**, within-session switching, where the participant would start focused and then explore from there. P19 described how in Switch they could say:

> 'I don't want to watch that. I'm only going to go in these directions.' So even though I spent less time, the majority of it felt very, very much like I wanted to watch the initial video that I watched, and I felt good about doing that.

Using an initial video to direct their exploration helped participants stick to their original goal for use and feel satisfied about it.

Focus better supported participants even when they spent the same amount of time in it as Explore because the quality of the content that they watched better aligned with their goals. The same participant described:

> [With Focus] it's easier to feel satiated: 'Okay, yeah, that was cool. I'm glad I watched that. I don't feel like I necessarily wasted my time'… And I think it's funny to see that I spent more time in the Focus, but the quality of it felt better. (P19)

In contrast to Explore, participants said Focus did not induce the same feelings of guilt about *"wasting time."*

6 DISCUSSION

Overall, in our survey we found that about half of the participants in our sample of general YouTube users were either actively or imminently trying to change something about their YouTube use. Reducing the quantity of time spent on YouTube was a common goal, but so was changing the quality of the content, for example by increasing consumption of educational videos or decreasing that of recommendations.

In our field experiment, Switch better supported sense of agency, satisfaction, and goal alignment than Explore alone, and also offered greater satisfaction than Focus alone. Focus alone was also rated significantly higher than Explore on all three of these measures. Unexpectedly, this was true despite the fact that time spent in Focus (median: 145 minutes) was actually somewhat higher than in Explore (median: 117). In terms of switching behavior, participants engaged in both between-session switching to stick to an original intention for that session of use (**Use Case 3**) and within-session switching to start focused and then explore from there (**Use Case 4**). When Explore was the only option it was rated relatively low in terms of user experience, but the high ratings for Switch and the prevalence of Use Cases 3 and 4 show that there were situations where users did find it beneficial. We discuss these findings and their design implications.

6.1 Quality of Time, Not Just Quantity

For the screen time tools that come pre-installed on every iOS and Android smartphone (Apple Screen Time and Google's Digital Wellbeing) and in the vast majority of third-party digital wellbeing apps and extensions [47], the goal is to reduce time spent on the phone or in certain apps. However, our survey results add to the growing chorus of digital wellbeing researchers calling for people to move beyond just screen time as the metric of success [11, 18, 45]. Specifically, in the case of YouTube, we find that decreasing quantity of time accounted for only 37% of the goals that survey participants had. A larger category of goals was shifting the quality of consumption (51% in total), either by increasing a certain type of content (e.g., educational, new, or meaningful videos) or decreasing another (e.g., meaningless, sponsored, or recommended videos). In other words, screen time did matter for participants, but it was a less common goal than changing the quality of time spent on YouTube, which is largely unsupported by current screen time tools.

Our results also contribute empirical evidence that designers can support concerned users and improve user experience without necessarily reducing screen time. To our surprise, SwitchTube study participants actually spent more time in Focus (median: 145 minutes) than in Explore (median: 117), and although many of them had a goal of reducing their time spent on YouTube, they still rated it significantly higher than Explore in terms of sense of agency, satisfaction, and goal alignment. Participants said that even if they spent the same amount of time in Focus as in Explore, that time was usually better spent. Similar results have recently been found for a version of Twitter that was redesigned to support a greater sense of agency [71]. The term "screen time" has provided a common vocabulary for many users to voice their concerns with digital device use, but it has also led to an overemphasis on *time* as the singular metric of success and limited the solutions imagined by design researchers and practitioners.

One key consideration in analyzing our experiment is that homepage recommendations were not personalized in SwitchTube as they are in the YouTube app. Since recommendations (both personalized and non-personalized) account for about 70% of watch time on YouTube [62], this means that the Explore Version may be difficult to compare against the current YouTube app. Adding personalized recommendations to SwitchTube would likely have increased the number of recommended videos watched and the amount of time spent in the Explore Version, but not in the Focus Version in which homepage recommendations were turned off by default. We expect that personalized recommendations would have reduced irrelevant recommendations but also increased temptation, perhaps leading Explore to be rated higher on satisfaction but lower on goal alignment—as we had originally hypothesized. Future work could investigate how personalized recommendations influence time spent and user experience ratings from the perspective of digital wellbeing.

6.2 Adaptable Commitment Interfaces for Digital Wellbeing

The term "adaptable commitment interface" sounds like a paradox. How can one "commit" if one can easily switch away from that commitment at any time without any cost? An ACI lacks the core mechanism behind commitment devices: that the present self imposes a meaningful cost (in terms of time, money, or effort) upon the future self for abandoning their original goal. Unlike the concept of the goal reminders that are also used in the digital wellbeing space [27, 48], an ACI does not necessarily provide informational support. Yet the ACI we designed (the Switch Version of SwitchTube) received higher participant ratings for sense of agency, satisfaction, and goal alignment than an interface that lacked commitments (the Explore Version of SwitchTube). The ACI also outperformed a non-adaptable commitment interface (the Focus Version of SwitchTube) in terms of participant ratings for satisfaction.

We believe that Switch worked as an ACI because it avoided triggers for distraction in the first place. This aligns with the emerging consensus in psychological research that people who are effective self-regulators cannot necessarily use willpower to resist temptations once encountered, but are instead skilled at avoiding them altogether [19, 20]. For digital wellbeing designers, this suggests that rather than waiting to intervene "just-in-time" by reacting to tempting content with a barrier or goal reminder, they should consider how to prevent that content from appearing in the first place. Of course, this could also be implemented by just entirely blocking or limiting access to an app with distracting content as is currently done in many screen time tools (a commitment interface), but the cost is that it can also indiscriminately block access to desired features and provoke psychological reactance in response to the restrictions it imposes on user autonomy [43, 46]. An advantage of ACIs is that they impose no such autonomy restriction: in Switch, if the user wanted to toggle over to Explore Mode, they were free to do so. Participants reported a high sense of agency in the Switch Version.

One clear design implication is that users would benefit from a "Focus Mode" even within an app such as YouTube that is commonly regarded as a service for entertainment. And of course, YouTube does provide entertainment, but users often have goals of using it for other purposes as well—for example, education both formal and informal, as in the SwitchTube participants who described watching videos to learn Spanish and to fix a flat tire. Rather than offering Focus Mode only in "productivity" tools like Microsoft Word (where there are far fewer distractions to begin with), there is a strong case for adding it to services that people rely upon for many different use cases, such as YouTube, Twitter, and Facebook.

Unfortunately our finding that a Focus Mode would benefit the user experience does not mean that social media companies will necessarily be keen adopt it. In fact, the design of social media seems to be trending in the opposite direction: short videos that play on continuous autoplay have powered the rise of TikTok and are now being emulated by YouTube, Facebook, and Instagram [50]. As a matter of practicality, designers needing to convince other interested parties that a Focus Mode will not necessarily harm the business interests of advertising-supported services might cite our result that the Focus Version did not reduce time spent in Switch-Tube.[6]

As a matter of principle though, designs that support user agency and wellbeing should be encouraged *even if* they go against the business interests (e.g., reduce time spent on a service). As such, our findings may also be of interest to policymakers and regulators who are tasked with crafting and implementing incentives for online platforms. For example, the Digital Services Act that was approved by the European Council in October 2022 calls for an end to dark patterns that *"distort and impair user autonomy, decision-making and choice"* [55, p.57]. Translating such an ambitious and sweeping directive into action will require further studies in the vein of SwitchTube that attempt to enact and assess abstract values such as 'user sense of agency' in digital interfaces.

6.3 From Adaptable to Adaptive Commitment Interfaces

SwitchTube is an *adaptable* commitment interface, but it also holds implications for *adaptive* commitment interfaces. In Switch, the user can manually toggle between two modes, one that lets the system take control (Explore Mode) and the other that helps hold the user to their intention (Focus Mode). Yet this manual labor does place a certain burden upon the user and might instead be automatically performed by the system itself, i.e., an adaptive commitment interface. For example, if YouTube has high confidence that the user is visiting with a specific intention, it could present a search-only interface and hide all recommendations. Conversely, if YouTube expects that the user is visiting without an intention in mind, it could take control as in its Leanback mode for "effortless viewing" that autoplays a never-ending stream of recommendations [24].

Already, machine-learning systems can predict the specificity of a user's intent for visiting a site with some accuracy [12, 29], opening the door for interfaces that adapt the level of sense of agency to meet a user's needs. Yet given that these predictions do still have a substantial error rate at present and erroneous predictions could provoke user reactance, what might a more reliable first step towards such an adaptive commitment interface look like?

Use Case 4, within-session switching, suggests one such direction. Participants started in Focus Mode to search and then switched to Explore Mode after they had performed their search to branch out from there. Because search queries are a strong signal of intent, this could be an ideal opportunity for the system to make a well-informed prediction of the user's needs and adapt the interface accordingly. For example, such a system would always start in the search-first Focus Mode. Searches such as "flat tire" that suggest a specific intention would keep the user in Focus Mode whereas a query such as "standup comedy" that suggests an openness to simply being entertained would automatically switch them to Explore Mode. Such a search-informed adaptive commitment interface could offer semi-automatic support for when a user has

---

[6]However, as discussed earlier, this result may have been different if the Explore Version had of included *personalized* video recommendations.

a focused goal, but still afford the flexibility to explore. In other words, the user provides the general direction of their intention and the system supports them to optimize for that intent, adapting into the interface for either focus or exploration.

6.4 Limitations and Future Work

A limitation of our work is that we compared the time spent in each of the three versions of SwitchTube, but did not measure what, if any, spillover effect there was on time spent on YouTube itself. Prior work has found that productivity interventions on one platform do not redistribute time to other apps or platforms [38]. In the case of SwitchTube though, it is possible that because Explore is the most similar version to the existing interface of YouTube it was seen as a poor substitute, whereas Focus and Switch offered something different, so were used more often as a substitute for YouTube use.

Our findings also highlight an important need for research on how to help users *shift* their use in desired directions, not just how to reduce it. Again due to limitations of the YouTube Data API, we had to make due with kludgy workarounds (e.g., appending "viral" to the search query) to change the quality of the content in our study app. Future work in digital wellbeing might devise clever ways to manipulate the inner workings of recommender systems and search algorithms that drive so much of social media use today. To inform the design of such algorithms, one starting point could be to analyze the content of the dataset of 2063 video play events in SwitchTube that we make available via the Open Science Framework: https://osf.io/z735n

7 CONCLUSION

Our work steps beyond the current paradigm of screen time tools and demonstrates how a digital wellbeing design might also support users in shifting the *quality* of their consumption. We designed and developed SwitchTube, an Android app with three different versions for watching YouTube videos, and deployed it in a field experiment with 46 US YouTube users over three weeks. Focus was rated higher than Explore on user experience metrics, but surprisingly it did not reduce the amount of time that participants spent watching videos. Yet the Switch Version, which provided the ability to toggle between both Explore and Focus Mode, was rated the highest of all. In other words, the Switch Version proved effective as an adaptable commitment interface (ACI) that provided the user with the flexibility to switch, but still helped them to follow their initial intention, providing the best of both worlds. Designers interested in digital wellbeing should consider implementing ACIs that support the user to explore when they want to explore and focus when they want to focus.

ACKNOWLEDGMENTS

REFERENCES

[1] Adam Alter. 2017. *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked*. Penguin, . https://market.android.com/details?id=book-uRyMDAAAQBAJ

[2] Janarthanan Balakrishnan and Mark D Griffiths. 2017. Social media addiction: What is the role of content in YouTube? *Journal of behavioral addictions* 6, 3 (Sept. 2017), 364–377. https://doi.org/10.1556/2006.6.2017.058

[3] Eric P S Baumer, Phil Adams, Vera D Khovanskaya, Tony C Liao, Madeline E Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, Leaving, and (Re)Lapsing: An Exploration of Facebook Non-use Practices and Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 3257–3266. https://doi.org/10.1145/2470654.2466446

[4] Eric P S Baumer, Rui Sun, and Peter Schaedler. 2018. Departing and Returning: Sense of Agency As an Organizing Concept for Understanding Social Media Non/Use Transitions. *Proc. ACM Hum. -Comput. Interact.* 2, CSCW (Nov. 2018), 23:1–23:19. https://doi.org/10.1145/3274292

[5] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Comput. Surv.* 50, 6 (Dec. 2017), 93:1–93:40. https://doi.org/10.1145/3123988

[6] Bianca Bosker. 2016. What Will Break People's Addictions to Their Phones? http://www.theatlantic.com/magazine/archive/2016/11/the-binge-breaker/501122/

[7] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (Aug. 2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

[8] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2018. Thematic Analysis. In *Handbook of Research Methods in Health Social Sciences*, Pranee Liamputtong (Ed.). Springer Singapore, Singapore, 1–18. https://doi.org/10.1007/978-981-10-2779-6_103-1

[9] Gharad Bryan, Dean Karlan, and Scott Nelson. 2010. Commitment Devices. *Annual review of economics* 2, 1 (Sept. 2010), 671–698. https://doi.org/10.1146/annurev.economics.102308.124324

[10] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[11] Marta E Cecchinato, John Rooksby, Alexis Hiniker, Sean Munson, Kai Lukoff, Luigina Ciolfi, Anja Thieme, and Daniel Harrison. 2019. Designing for Digital Wellbeing: A Research & Practice Agenda. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19, Paper W17)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3298998

[12] Justin Cheng, Caroline Lo, and Jure Leskovec. 2017. Predicting Intent Using Activity Logs: How Goal Specificity and Temporal Range Affect User Behavior. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 593–601. https://doi.org/10.1145/3041021.3054198

[13] Hyunsung Cho, Daeun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. 2021. Reflect, not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 1–36. https://doi.org/10.1145/3479600

[14] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18, Paper 141)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173715

[15] Emily I M Collins, Anna L Cox, Jon Bird, and Daniel Harrison. 2014. Social networking use and RescueTime: The issue of engagement. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. Association for Computing Machinery, New York, NY, United States, 687–690. https://doi.org/10.1145/2638728.2641322

[16] Anna L Cox, Sandy J J Gould, Marta E Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. ACM, New York, NY, USA, 1389–1397. https://doi.org/10.1145/2851581.2892410

[17] Liam Delaney and Leonhard K Lades. 2017. Present bias and everyday self-control failures: a day reconstruction study. *Journal of behavioral decision making* 30, 5 (2017), 1157–1167. https://doi.org/10.1002/bdm.2031

[18] M A Devito, A M Walker, J Birnholtz, K Ringland, and others. 2019. Social Technologies for Digital Wellbeing Among Marginalized Communities. *Publication of the 2019 …* (2019), 449–454. https://dl.acm.org/doi/abs/10.1145/3311957.3359442

[19] Angela L Duckworth and James J Gross. 2020. Behavior Change. *Organizational behavior and human decision processes* 161, Suppl (Nov. 2020), 39–49. https://doi.org/10.1016/j.obhdp.2020.09.002

[20] Angela L Duckworth, Rachel E White, Alyssa J Matteucci, Annie Shearer, and James J Gross. 2016. A Stitch in Time: Strategic Self-Control in High School and

College Students. *Journal of educational psychology* 108, 3 (April 2016), 329–341. https://doi.org/10.1037/edu0000062

[21] Sindhu Kiranmai Ernala, Moira Burke, Alex Leavitt, and Nicole B Ellison. 2020. How Well Do People Report Time Spent on Facebook? An Evaluation of Established Survey Questions with Recommendations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376435

[22] James Fogarty. 2017. Code and Contribution in Interactive Systems Research. In *ACM CHI Conference 2017. Workshop on #HCI.Tools: Strategies and Best Practices for Designing, Evaluating, and Sharing Technical HCI Toolkits.* washington.edu, washington, 1–4. https://homes.cs.washington.edu/~jfogarty/publications/workshop-chi2017-codeandcontribution.pdf

[23] Mozilla Foundation. 2020. YouTube Regrets — Mozilla Foundation. Retrieved May 1, 2020 from https://foundation.mozilla.org/en/youtube/regrets/

[24] Google. . YouTube Leanback offers effortless viewing. Retrieved September 12, 2020 from https://youtube.googleblog.com/2010/07/youtube-leanback-offers-effortless.html

[25] Marco Gui and Luca Stanca. 2009. *Television viewing, satisfaction and happiness: Facts and fiction.* Technical Report. Università degli Studi di Milano - Bicocca. http://repec.dems.unimib.it/repec/pdf/mibwpaper167.pdf

[26] Francesca Gullà, Lorenzo Cavalieri, Silvia Ceccacci, Michele Germani, and Roberta Bevilacqua. 2015. Method to Design Adaptable and Adaptive User Interfaces. In *HCI International 2015 - Posters' Extended Abstracts.* Springer International Publishing, Los Angeles, CA, USA, 19–24. https://doi.org/10.1007/978-3-319-21380-4_4

[27] Alexis Hiniker, Sungsoo (ray) Hong, Tadayoshi Kohno, and Julie A Kientz. 2016. MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, United States, 4746–4757. https://doi.org/10.1145/2858036.2858403

[28] Alexis Hiniker, Bongshin Lee, Kiley Sobel, and Eun Kyoung Choe. 2017. Plan & Play: Supporting Intentional Media Use in Early Childhood. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) *(IDC '17)*. Association for Computing Machinery, New York, NY, USA, 85–95. https://doi.org/10.1145/3078072.3079752

[29] Alexis Hiniker, Shwetak N Patel, Tadayoshi Kohno, and Julie A Kientz. 2016. Why would you do that? predicting the uses and gratifications behind smartphone-usage behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* dl.acm.org, New York, NY, United States, 634–645. http://dl.acm.org/citation.cfm?id=2971762

[30] Wilhelm Hofmann, Roy F Baumeister, Georg Förster, and Kathleen D Vohs. 2012. Everyday temptations: an experience sampling study of desire, conflict, and self-control. *Journal of personality and social psychology* 102, 6 (June 2012), 1318–1335. https://doi.org/10.1037/a0026545

[31] Daniel Kahneman. 2011. *Thinking, Fast and Slow.* Farrar, Straus and Giroux, . https://books.google.com/books?id=SHvzzuCnuv8C

[32] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1 (March 2019), 29. https://doi.org/10.1145/3314403

[33] Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType: Lockout Task Intervention for Discouraging Smartphone App Use,". In *ACM CHI.* Association for Computing Machinery, New York, NY, United States, 1–12. https://doi.org/10.1145/3290605.3300927

[34] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, Kwonhyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 272–283. https://doi.org/10.1145/2858036.2858428

[35] Jane E Klobas, Tanya J McGill, Sedigheh Moghavvemi, and Tanuosha Paramanathan. 2019. Problematic and extensive YouTube use: first hand reports. *Online Information Review* 43, 2 (Jan. 2019), 265–282. https://doi.org/10.1108/OIR-01-2018-0032

[36] Minsam Ko, Seungwoo Choi, Koji Yatani, and Uichin Lee. 2016. Lock n'LoL: group-based limiting assistance app to mitigate smartphone distractions in group activities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* dl.acm.org, New York, NY, United States, 998–1010. https://dl.acm.org/citation.cfm?id=2858568

[37] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: A Group-based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. ACM, New York, NY, USA, 1235–1245. https://doi.org/10.1145/2675133.2675244

[38] Geza Kovacs, Drew Mylander Gregory, Zilin Ma, Zhengxuan Wu, Golrokh Emami, Jacob Ray, and Michael S Bernstein. 2019. Conservation of Procrastination: Do

[39] Productivity Interventions Save Time Or Just Redistribute It?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 330:1–330:12. https://doi.org/10.1145/3290605.3300560

[39] Geza Kovacs, Zhengxuan Wu, and Michael S Bernstein. 2021. Not Now, Ask Later: Users Weaken Their Behavior Change Regimen Over Time, But Expect To Re-Strengthen It Imminently. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445695

[40] J R Landis and G G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (March 1977), 159–174. https://www.ncbi.nlm.nih.gov/pubmed/843571

[41] Simone Lanette, Phoebe K Chua, Gillian Hayes, and Melissa Mazmanian. 2018. How Much is 'Too Much'?: The Role of a Smartphone Addiction Narrative in Individuals' Experience of Use. *Proc. ACM Hum. -Comput. Interact.* 2, CSCW (Nov. 2018), 101:1–101:22. https://doi.org/10.1145/3274370

[42] Sara Lebow. 2022. TikTok to surpass YouTube in US—and come after other apps in the process. https://www.emarketer.com/content/tiktok-surpass-youtube. https://www.emarketer.com/content/tiktok-surpass-youtube Accessed: 2022-5-17.

[43] Kai Lukoff, Ulrik Lyngs, and Lize Alberts. 2022. Designing to Support Autonomy and Reduce Psychological Reactance in Digital Self-Control Tools. In *Position Papers for the Workshop "Self-Determination Theory in HCI: Shaping a Research Agenda" at the Conference on Human Factors in Computing Systems (CHI '22).* ACM, New Orleans, LA, USA, 5.

[44] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3411764.3445467

[45] Kai Lukoff and Sean Munson. 2019. Digital wellbeing is way more than just reducing screen time. https://uxdesign.cc/digital-wellbeing-more-than-just-reducing-screen-time-46223db9f057. https://uxdesign.cc/digital-wellbeing-more-than-just-reducing-screen-time-46223db9f057 Accessed: 2020-10-1.

[46] Ulrik Lyngs, Kai Lukoff, Laura Csuka, Petr Slovák, Max Van Kleek, and Nigel Shadbolt. 2022. The Goldilocks Level of Support: Using User Reviews, Ratings, and Installation Numbers to Investigate Digital Self-Control Tools. *International journal of human-computer studies* (June 2022), 102869. https://doi.org/10.1016/j.ijhcs.2022.102869

[47] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 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 - CHI '19.* ACM Press, Glasgow, Scotland Uk, 1–18. https://doi.org/10.1145/3290605.3300361

[48] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 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.* Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376672

[49] Lucas Matney. 2017. YouTube has 1.5 billion logged-in monthly users watching a ton of mobile video. *TechCrunch* (June 2017), . http://techcrunch.com/2017/06/22/youtube-has-1-5-billion-logged-in-monthly-users-watching-a-ton-of-mobile-video/

[50] Megan McCluskey. 2022. YouTube's Next Move in the Race to Beat TikTok. *Time* (Sept. 2022), . https://time.com/6214679/youtube-shorts-tiktok-creators/

[51] Alberto Monge Roffarello and Luigi De Russis. 2022. Towards Understanding the Dark Patterns That Steal Our Attention. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (New Orleans, LA, USA) *(CHI EA '22, Article 274)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3491101.3519829

[52] Alberto Monge Roffarello, Kai Lukoff, and Luigi De Russis. 2023. Defining and Identifying Attention Capture Damaging Patterns in Digital Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, New York, NY, USA. https://doi.org/10.1145/3544548.3580729

[53] MrBeast. 2021. If You Can Carry $1,000,000 You Keep It! https://www.youtube.com/watch?v=ktyJIj6i4Qw

[54] John C Norcross, Paul M Krebs, and James O Prochaska. 2011. Stages of change. *Journal of clinical psychology* 67, 2 (Feb. 2011), 143–154. https://doi.org/10.1002/jclp.20758

[55] European Parliament. 2022. Digital Services Act. https://www.europarl.europa.eu/doceo/document/TC1-COD-2020-0361_EN.pdf

[56] Ethan Plaut. 2014. *Commitments Not to Communicate Before and After Digital Media: A Study of the Will and Ways to Disconnect Under Changing Conditions.* Ph. D. Dissertation. Stanford University.

[57] James O Prochaska. 2013. Transtheoretical Model of Behavior Change. In *Encyclopedia of Behavioral Medicine*, Marc D Gellman and J Rick Turner (Eds.).

Springer New York, New York, NY, 1997–2000. https://doi.org/10.1007/978-1-4419-1005-9_70

[58] SarahBethYoga. 2019. 15 minute Hands Free Morning Yoga Routine | Sarah Beth Yoga. https://www.youtube.com/watch?v=t3nzh01axMI

[59] SarahBethYoga. 2019. 15 minute Morning Yoga Routine | FULL BODY POWER YOGA FLOW. https://www.youtube.com/watch?v=HgxBGC7OgJA

[60] Magy Seif El-Nasr, Shree Durga, Mariya Shiyko, and Carmen Sceppa. 2015. Data-Driven Retrospective Interviewing (DDRI): A proposed methodology for formative evaluation of pervasive games. *Entertainment computing* 11 (Nov. 2015), 1–19. https://doi.org/10.1016/j.entcom.2015.07.002

[61] Joan E Solsman. 2018. Ever get caught in an unexpected hourlong YouTube binge? Thank YouTube AI for that. https://www.cnet.com/news/youtube-ces-2018-neal-mohan/. https://www.cnet.com/news/youtube-ces-2018-neal-mohan/ Accessed: 2020-5-1.

[62] Joan E Solsman. 2018. Ever get caught in an unexpected hourlong YouTube binge? Thank YouTube AI for that. https://www.cnet.com/news/youtube-ces-2018-neal-mohan/. https://www.cnet.com/news/youtube-ces-2018-neal-mohan/ Accessed: 2020-5-1.

[63] Jonathan A Tran, Katie S Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300542

[64] United States Census Bureau. . QuickFacts: United States. https://www.census.gov/quickfacts/fact/table/US/PST045219

[65] James Williams. 2018. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge University Press, United Kingdom.

[66] Kaitlin Woolley and Marissa A Sharif. 2022. Down a Rabbit Hole: How Prior Media Consumption Shapes Subsequent Media Consumption. *JMR, Journal of marketing research* 59 (2022), 453—-471. https://journals.sagepub.com/doi/abs/10.1177/00222437211055403

[67] Janine L Wright, Jillian L Sherriff, Satvinder S Dhaliwal, and John C L Mamo. 2011. Tailored, iterative, printed dietary feedback is as effective as group education in improving dietary behaviours: results from a randomised control trial in middle-aged adults with cardiovascular risk factors. *The international journal of behavioral nutrition and physical activity* 8, 1 (May 2011), 43. https://doi.org/10.1186/1479-5868-8-43

[68] Tim Wu. 2016. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Vintage, New York.

[69] Xuhai Xu, Tianyuan Zou, Han Xiao, Yanzhang Li, Ruolin Wang, Tianyi Yuan, Yuntao Wang, Yuanchun Shi, Jennifer Mankoff, and Anind K Dey. 2022. TypeOut: Leveraging Just-in-Time Self-Affirmation for Smartphone Overuse Reduction. In *CHI Conference on Human Factors in Computing Systems (CHI '22, Article 442)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3491102.3517476

[70] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark Patterns in the Design of Games. In *Foundations of Digital Games 2013*. diva-portal.org, Interactive Institute. http://www.diva-portal.org/smash/record.jsf?pid=diva2:1043332

[71] Mingrui Ray Zhang, Kai Lukoff, Raveena Rao, Amanda Baughan, and Alexis Hiniker. 2022. Monitoring Screen Time or Redesigning It? Two Approaches to Supporting Intentional Social Media Use. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22, Article 60)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3491102.3517722

[72] Shoshana Zuboff. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books, New York.