Designing to Support Sense of Agency for Time Spent on Digital Interfaces

Kai Lukoff

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Sean A. Munson, Co-Chair

Alexis Hiniker, Co-Chair

Eric P. S. Baumer

Program Authorized to Offer Degree:

Human Centered Design & Engineering

University of Washington

Abstract

Designing to Support Sense of Agency for Time Spent on Digital Interfaces

Kai Lukoff

Co-Chairs of the Supervisory Committee:

Sean A. Munson
Human Centered Design & Engineering

Alexis Hiniker
Information School

App designers often exploit psychological vulnerabilities to maximize clicks, views, and time on site. When people attempt to resist such media use, their failure rate is higher than for any other temptation in everyday life. Consequently, users often report feeling dissatisfied and regretful of the time that they spend in apps. In response, concerned design practitioners and researchers have innovated 'screen time tools' that let users track and limit the time they spend on digital devices. Yet users report that reducing screen time is a poor proxy for their actual goals, that they are concerned with not only the quantity but also the quality of the time they spend online, so the problem persists.

In this dissertation, I investigate how to respect the user's time and attention by designing digital interfaces for a greater sense of user agency, i.e., the experience of control over one's actions and their outcomes. My research on the YouTube mobile app, a common site of problematic use, finds that a majority of user goals are about shifting the quality of the content they consume on smartphones, not the quantity. Through a survey and co-design activities, I identify specific features that lead users to feel more or less control over how they spend their time on YouTube. Based on these features, I design and develop 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 field deployment of the SwitchTube app with 46 U.S. participants, I find that Focus Mode helps them realize a greater sense of agency *without* reducing their time spent in the app.

My work highlights the need to think beyond 'screen time' and advances sense of agency as an alternative lens for addressing user frustrations. I highlight how the design community might identify and call out 'attention capture dark patterns,' conceptualize and measure sense of agency, and how flexible interfaces might adapt support for sense of agency to suit different use cases. Ultimately, sense of agency is not only associated with positive technology use outcomes, but also matters to users in its own right as a basic psychological need.

Dissertation Committee

1. Sean A. Munson, University of Washington, HCDE (co-chair)
2. Alexis Hiniker, University of Washington, Information School & HCDE Affiliate (co-chair)
3. Eric P. S. Baumer, Lehigh University, CSE (reading committee member)
4. Julie Kientz, University of Washington, HCDE (general committee member)
5. Ryan Calo, University of Washington, School of Law (graduate school representative)

Table of Contents

Acknowledgements 7

Glossary 11

Thesis Statements 12

CHAPTER 1 — INTRODUCTION 13

CHAPTER 2 — DUAL SYSTEMS THEORY 17

2.1 — Behaviorism: Users as Skinner's Pigeons 17

2.2 — Rational Choice Theory: Users as Rational Actors 19

2.3 — Dual Systems Theory: A Model of Cognition 21

2.4 — An Integrative Dual Systems Model 22

2.5 — Early Self-Regulation Interventions are More Effective 25

CHAPTER 3 — NUDGING, DARK PATTERNS, AND USER FRUSTRATIONS 27

3.1 — Nudge Theory 28

3.2 — Dark Patterns 28

3.3 — User Frustrations 30

CHAPTER 4 - DIGITAL WELLBEING TOOLS 33

4.1 — Reflective System Designs 33

4.2 — Automatic System Designs 34

CHAPTER 5 — UNDERSTANDING HOW APP USE LEADS TO A LOSS OF SENSE OF AGENCY FOR TIME SPENT ON DIGITAL INTERFACES 37

5.1 — Background and Motivation 38

5.2 — Methods 42

5.3 — Results 49

5.4 — Discussion 66

5.5 — Limitations and Future Work 71

CHAPTER 6 — DESIGNING TO INFLUENCE SENSE OF AGENCY OVER APP USE 73

6.1 — Background and Motivation 76

6.2 — Study 1: Survey of 120 YouTube Users 79

6.3 — Study 2: Co-design with YouTube Users 91

6.4 — Overall Discussion 100

CHAPTER 7 — SWITCHTUBE: TOGGLING BETWEEN EXPLORE AND FOCUS MODE WHEN WATCHING YOUTUBE VIDEOS 109

7.1 — Background and Motivation 109

7.2 — Survey: Stages of Change for YouTube Users 115

7.3 — SwitchTube Field Experiment: Evaluating an Adaptable Commitment Interface 121

7.4 — Results and Analysis 135

7.4 — Discussion 148

CHAPTER 8 — REFLECTING ON SENSE OF AGENCY AS A NEW LENS FOR DIGITAL WELLBEING 154

8.1 — The Need for Sense of Agency 154

8.2 — The Challenge of False Sense of Agency 155

8.3 — The Challenge of Measurement 156

8.4 — The Challenge of Time-Inconsistent Preferences 158

CHAPTER 9 — FUTURE WORK AND IMPACT: INFRASTRUCTURAL, SOCIAL, AND SYSTEMIC CONSIDERATIONS 162

9.1 — Research Infrastructure for Systematic App Reviews 162

9.2 — Social Considerations 164

9.3 — Systemic Considerations 165

Conclusion 169

References 170

Acknowledgements

In search of autonomy. In the fall of 2015, when I applied to the Ph.D. program in Human-Centered Design & Engineering at the University of Washington I wrote that I wanted "greater autonomy" to pursue topics that I found meaningful as compared to the work that I had been doing in industry. Little did I know that sense of agency and meaning would themselves become the focus of my dissertation research. This would never have been possible without the support I've received from advisors, mentors, colleagues, friends, and family who not only respected my own need for autonomy, but also encouraged me to pursue it as a topic of research.

Of course, autonomy does not entail loneliness, and it certainly does not entail doing all of the work on one's own. Advisors, mentors, teachers, students, friends, and family have all offered me kindness, encouragement, and thoughtfulness at every step of the way. It has been a joy and a privilege to receive your support, and I will try to pass it on to others as I move forward in my own career.

To Sean Munson. For modeling Professoring. For providing thoughtful and thorough advice on everything from experimental design to conference reimbursements. For diligently reading and editing all of my publications. For finding funding for this research, even at times when I was exploring new domains or caring for my daughter. For always encouraging reflection, sometimes nudging, and never shoving.

To Alexis Hiniker. For generously advising me when you were still a Ph.D. student yourself, and helping me find my way ever since. For snapping me out of 'frozen' Zoom calls with your creativity and pragmatism. For endlessly debugging the SwitchTube app and all of my projects in general. For also working tirelessly to secure funding to support this line of work. And for the warm welcome into Ph.D. parenthood!

And to all of my dissertation committee members for their engagement with my dissertation research and support for my career more broadly. **To Julie Kientz** for teaching me how to teach and be present for students, for modeling how to balance competing

priorities as a researcher, teacher, and parent, and for encouraging me to examine not just the meaningless experiences with technology, but also the meaningful ones. **To Eric Baumer** for responding to my random email at my first ever CHI in Denver and supporting me ever since. You have always engaged with my work with great thoughtfulness and curiosity and given me so much to reflect on. **To Ryan Calo** for enthusiastically disregarding traditional departmental boundaries, for prompting me to rigorously consider various definitions of attentional harms, and for helping promote my work to a wider audience.

To all of the faculty and colleagues at UW who have made my time here so memorable and productive, in particular: David Ribes, Gary Hsieh, Jennifer Turns, David McDonald, Mako Hill, Tadayoshi Kohno, Elena Agapie, Ravi Kakar, Daniel Epstein, Lucas Colusso, Christina Chung, Mia Suh, Tom Wilson, Himanshu Zade, Mingrui (Ray) Zhang, Amanda Baughan, Karina Shirokova, Raveena Rao, Larry Tian, Cissy Yu, James Choi, Kaiyue Fan, Vera Liao, Stefania Gueroguieva, Erika Dillman, Taoxi Li, and Yuan Zhuang. And to all of the students who have participated in the nine directed research groups (DRGs) I have run during my time at HCDE: your skills, enthusiasm, and creativity have astonished me and improved the work immeasurably.

To the mentors, colleagues, and collaborators from beyond UW who have opened up the worlds of academia and industry to me, in particular: Paul Resnick, Sarita Schoenebeck, Carol Moser, Petr Slovak, Max van Kleek, Jun Zhao, Jofish Kaye, Janice Tsai, Brian Y. Lim, Marta Ceccinato, Colin Gray, David Victorson, Carly Maletich, Bill Bishop, Kaiser Kuo, Yaping Liu, and Junyu Wang.

To Ulrik Lyngs. Without you, I'm not sure I would have survived the slow grind of a Ph.D. and the work certainly would not have been half as good. As part of my general exam, I read Michael Lewis's *The Undoing Project*, the story of Daniel Kahneman and Amos Tversky's friendship and decades-long intellectual collaboration. Ever since, I have always thought of our relationship as a mini-version of that great partnership, hopefully without the acrimonious breakup and regrettably without the Nobel Prize—yet!

To all the inspirational teachers from my earlier years, from elementary school through college: Lisa Damico, Tim May, Carol Baucom, Joyce Barthman, Mary Jackson, Kylee Cecchini, Bobbin Tobin, Larry Ellison, Marty Hogue, Andrew Mitchell, James Ashley Morrison, David Kennedy, Kenneth Schultz, Terry Karl, Coit D. Blacker, and so many more.

To my dearest friends for listening to my complaints about Ph.D. student life, buying me beers, and serving as a sounding board for explaining my work outside of academia, in particular: Evan Perkiss, Bridget Hudson, Jacob Becklund, Daniel Miller, Hilary Wurr, James Pennington, Max Weiland, Phillip Schwarzhaupt, Sami Hamid, Chloé Kayser, Josh Harder, David Houska, Nelson Ray, and Micah Cratty.

To Maya, Sebastian, and Solia for feeding and hosting me on the farm at Tempelhof at the start of my academic journey and almost every year since. You have always welcomed me into your life with open arms even when our worlds sometimes feel so far apart, and reminded me that there is so much more in the world than technology, academia, and big cities.

To my father. I watched you write countless articles of your own late at night in the garden office and on long camper van journeys through the American West. And from an early age, you always pressured me—and everyone who did not immediately tell you to shut up—to *publish*, *publish*, *publish*. Thankfully, for every one part pressure, your recipe called for two parts support. How many papers of mine have you edited throughout the years? I never set out to make you proud, but somehow I did? I hope I have.

To my mother. You *never* pressured me to publish. Instead, you always encouraged me to do meaningful things and modeled how to do so through your own work caring for the terminally ill and their families. When you returned from work at the end of the day physically and emotionally drained, I could see how much the work meant to you and to the people you cared for. And then you still cared for us too.

Für Mummi. Du hast mir immer ermutigt, meine Arbeit jemandem zu erklären, der nichts von Technologie versteht, obwohl du besser verstehsts als jeder andere 100-Jährige auf diesem Planeten. Unsere regelmäßigen FaceTime-Anrufe waren eine wichtige Quelle der

Unterstützung für mich, genauso wie ich hoffe, dass sie es für dir waren. Du bist wohl froh, dass ich endlich kein Student mehr bin, sondern sogar ein "Dr." genau wie du. Naja, nicht ganz so wie du, besser Hanne anrufen.

To Satoko. In my study in Singapore, I examined how and when family members gave each other tangible, informational, and emotional support for healthy eating. You have given me all, always, and for everything. Including, most recently, helping me add citations during the last minute crunch for my first submission of the dissertation—I promise never to ask you to do that again. Like the birds currently gathering twigs for their nest on our front porch, I hope we build our own new nest in California.

To Saya. Your arrival has meant the world to me. I have shaken *the bell*, dream fed you every night at 2am, and changed about 1,500 diapers—and counting. And yet you have also supported me more than you can ever know. Every morning when you wake me up at 6am, I know that in the evening we will kitchen dance together to *One Direction*, *Elton John*, and *ABBA*—bouncing, clapping, and waving our arms wildly in the air. You have the moves that melt daddy's heart. No pressure to publish just yet, you still have a couple of years.

Glossary

The terminology around agency can be confusing. This glossary defines key terms as they are used in my dissertation.

1. Agency - Control over one's actions and their consequences [164]
    - Sense of agency - the feeling of control over one's actions and their outcomes (Borhani et al., 2017).
        - Feelings of agency - In-the-moment experience of control
        - Judgments of agency - Post-hoc attribution of control to the self or other. In this dissertation, I focus on this notion of sense of agency.
2. Control - A synonym for agency
    - I use the language of control to communicate my work to a broader public and study participants. For example, *"What are 3 things about the mobile app that lead you to feel least in control over how you spend your time on YouTube?"*
3. Autonomy - The capacity to take informed, uncoerced actions [211]
    - There is significant overlap between agency and autonomy. However, agency is more akin to a state of being that comes or goes, whereas autonomy is more like a trait that is stable over time, e.g., an autonomous country.
4. Self-regulation - Controlling one's behavior, emotions, and thoughts in the pursuit of long-term goals [101]
5. Self-control - The ability to override unwanted impulses [209]
    - Synonymous with willpower
    - Self-regulation vs. self-control: "*Self-regulation* is commonly used as an umbrella term for all regulatory processes in the service of goal-directed behavior, including automatic habits, and *self-control* more restrictively about conscious, deliberate attempts at overriding immediate impulses that conflict with one's goals" [143]

Thesis Statements

Users often experience a loss of agency when using smartphone apps. Although screen time tools are intended to help users manage their time on such apps, they often indiscriminately block access to beneficial features. In this dissertation, I propose and evaluate design approaches that are more targeted than the current screen time paradigm. I make three main claims:

TS1. The screen time tools currently pre-installed on smartphones provide insufficient control for users to target the specific aspects of apps that they find problematic

TS2. Specific design features within apps have predictable effects on sense of agency

TS3. Adaptable interfaces can support the user to choose the right level of sense of agency for their use case

CHAPTER 1 — INTRODUCTION

Smartphones are omnipresent in the modern environment. This offers great benefits in the form of ubiquitous and personalized access to information, communication, and entertainment. However, this constant access also entails constant temptations for distraction.

Of course, temptations, i.e, desires that conflict with their long-term goals, are common in everyday life. When people attempt to resist such temptations, their failure rate is 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, which people fail to resist 76% of the time according to the results of a day reconstruction study [59] (**Table 1**). In a similar study using the experience sampling method, media use again is again associated with a far higher failure rate than most other temptations [98]. Today, smartphones account for a steadily increasing share of this media use.

Type of temptation	Self-control failure rate
Using media	76%
Postponing a task	58%
Spending money	35%
Smoking	30%
Drinking alcohol	30%
Sports	27%
Eating	26%
Having leisure time	21%
Satisfying sexual desire	18%
Sleeping	14%

Table 1. Media use has the highest self-control failure rate of all temptations. Data from 2059 samples from 142 participants in a day reconstruction study [59].

Although the attention economy around media use arose over a century ago with newspapers and radio [244], smartphone apps dramatically enhance the arsenal of attention merchants. First, because of the high rate of smartphone ownership (e.g., 85% of adults in the US [253] and the close proximity of these devices (within the same room as the user almost 90% of the time [61], apps can capture attention at nearly any time and any place. Second, apps can target highly specific demographics (e.g., female MMORPG gamers in New York City) that would have been difficult to reach in the past. Algorithms can leverage this granular tracking data to target consumer frailty at an individual level [32]. The combination of omnipresence and algorithmic targeting of individuals make mobile apps a potent tool for attention merchants.

The result is that digital media use is associated with a number of *attentional harms*—from distractions when driving and socializing to a loss of productivity and sleep [35]. Individually these short-term costs may seem small, but collectively they undermine people's ability to spend their time in accordance with their values [240]. It follows that users often report feeling dissatisfied and regretful of the time that they spend on mobile apps [44,233]. For social media users, these frustrations often manifest themselves as 'lagging resistance': a desire to quit, but not doing so just yet [13].

One lens for understanding these user frustrations is sense of agency. Sense of agency can be defined as the feeling of control over one's actions and their outcomes [22]. Baumer et al. [15] highlight three key attributes that characterize sense of agency: (1) the feeling of awareness, i.e., that actions occur within the field of consciousness; (2) the feeling of intentional action, i.e., that actions originate from one's own beliefs and desires; and (3) the feeling of capacity to act, i.e., belief in one's ability to influence the environment. They find that a desire to increase one's sense of agency drives both decisions to engage *and* to disengage with social media. In this dissertation, I further develop this concept of sense of agency as a lens for understanding user frustrations and extend it to guide the design and evaluation of mobile apps.

Table 2 provides an overview of the chapters in this dissertation. In this chapter (Chapter 1), I introduce how sense of agency lies at the heart of digital wellbeing concerns. Chapter 2 reviews dual systems theory as a model for understanding why there is often a gap between user intentions and behaviors with regards to technology use. Chapter 3 examines how nudging and dark patterns lead to user frustrations with time spent on technology. Chapter 4 reviews existing digital wellbeing tools and the limitations of the current 'screen time' paradigm.

Table 2. Overview of Dissertation Chapters

Chapter title	Content	Contributions
Chapter 1 — Introduction	Motivation for my work, summary of my contributions, outline of my dissertation	
Chapter 2 — Dual Systems Theory	Dual systems theory as a model for understanding user psychology	Accessible synthesis of dual systems theory for the HCI community
Chapter 3 — Nudging, Dark Patterns, and User Frustrations	Nudging and dark patterns and the regret and lagging resistance that prompt technology non-use	
Chapter 4 — Digital Wellbeing Tools	A review of existing tools in industry and research, limitations of the current 'screen time' paradigm	
Chapter 5 — Understanding How App Use Leads to a Loss of Sense of Agency	Experience sampling study with 45 smartphone users & interviews with 11 of them	Understanding of app use associated with meaninglessness and a loss of control
Chapter 6 — Designing to Influence Agency Over App Use	Survey of 120 YouTube mobile app users and co-design with 13 YouTube mobile app users	Identification of design patterns that support or undermine sense of agency over time spent in mobile apps
Chapter 7 — SwitchTube: An Adaptable Commitment Interface for Watching YouTube Videos	Within-subjects field evaluation of the SwitchTube app with 46 participants & interviews with 16 of them	• Understanding how SwitchTube affects user experience • Lessons for the design of adaptable commitment interfaces
Chapter 8 — Reflecting on User Agency as Lens for Digital Wellbeing	Reflections on the need for sense of agency and challenges in measurement and temporality	Guidance for using sense of agency as a lens for understanding and designing for digital wellbeing
Chapter 9 — Future Work and Impact: Infrastructural, Social, and Systemic Considerations	Future opportunities for research infrastructure and impact via social and systemic interventions	

Chapters 5 to 7 follow the human-centered design process (**Figure 1**). I start by understanding the user, then move on to designing, and finally build and evaluate a design in-the-wild. Each phase asks a different Thesis Question (TQ):

- **TQ1:** What app use makes people feel a loss of agency? (Chapter 5)
- **TQ2:** What design features within apps can we change to support user sense of agency? (Chapter 6)
- **TQ3:** Do alternative versions of these apps support user sense of agency in-the-wild? (Chapter 7)

In actuality, this research process was messy and iterative. For example, the design phase also contributed to new user understanding, which in turn motivated new designs. But for the sake of clarity I tell the story in a more linear fashion.

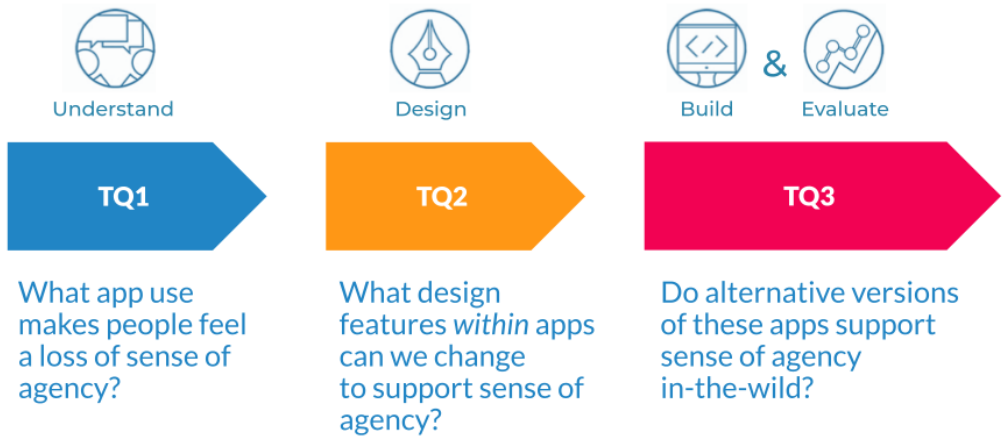


Figure 1. Chapters 4-6 of my dissertation follow the human-centered design process of understanding → designing → building & evaluating.

Chapter 8 reflects on the value of sense of agency as a design goal and also some of the challenges it faces: false sense of agency, measurement, and time-inconsistent preferences. Chapter 9 highlights opportunities for future work and impact, including research infrastructure for better understanding the mobile app ecosystem and how to support sense of agency over technology use via social and systemic interventions. Taken as a whole, my work provides both a conceptual lens and specific intervention ideas for individuals, designers, and policymakers to support a greater sense of agency over technology use.

CHAPTER 2 — DUAL SYSTEMS THEORY

In *Designing with the Mind in Mind*, Jeff Johnson describes how a core tenant of human-computer interaction (HCI) is that a good understanding of user psychology helps in designing good user interfaces [103]. In this chapter, I discuss three theories of the user that can inform design. The first two, behaviorism and rational choice theory, are presented to make explicit and criticize the assumptions behind the design of many apps that are designed for attention capture. The third, dual systems theory, is synthesized and discussed as a grounding for understanding existing user behaviors and developing new designs.

2.1 — Behaviorism: Users as Skinner's Pigeons

Behaviorism arose throughout the early- and mid-twentieth century in reaction to depth psychology (e.g., Freudian psychoanalysis), which analyzed the human psyche using introspection and therapeutic dialogue [90]. Behaviorists argued that these methods of analysis were too interpretive and unreliable. Instead, they championed laboratory experiments that measured precisely how animal and human behavior changed in response to stimuli in their environment. A classic example is B.F. Skinner's experiments, in which rats and pigeons pressed levers and received rewards such as food on different schedules [69]. One key insight is that a researcher (or designer) can train subjects (or users) to respond to an arbitrary stimulus (e.g., a lever or a notification icon) by associating it with a reward.

Behaviorism is relevant to designing technology for a sense of agency in at least two ways. First, it describes how a stimulus leads to behavior when reinforced by a reward. For instance, the knowledge that rewards on an intermittent schedule prompt more behavior than ones on a regular schedule, helps explain why one should expect users to visit Facebook more often when rewarding content appears at random times as opposed to batched together every four hours. The model of (1) stimulus → (2) behavior

→ (3) intermittent reward is a simple and powerful model for designing stimuli and rewards to maximize a desired behavior.

Second, behaviorism is important because the theory or adaptations thereof are highly influential among technology designers today. The Fogg Behavioral Model [70] and its popular translation, Nir Eyal's Hooked [67], both draw heavily upon behaviorism: the latter rephrases and extends model as (1) trigger → (2) action → (3) variable reward → (4) investment (**Figure 2**). The added "investment" phase is getting the user to spend resources (e.g., time, money, social proof, or create content) in the product. Here the understanding of user psychology is focused not on creating a "good" interface in an ethical sense or from the perspective of the user, but rather on yielding good investment returns for the designer/company.

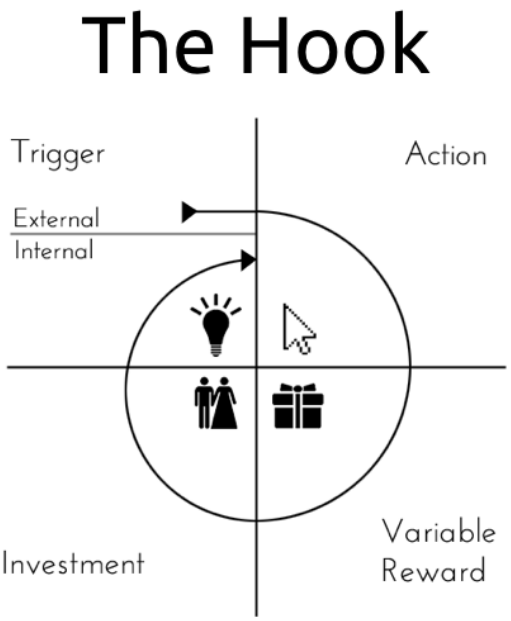


Figure 2. Nir Eyal's hooked model draws heavily upon behaviorism (trigger → action → variable reward) to get the user to 'invest' resources (e.g., time, money, social proof) in the product. Figure adapted from [67] by [51]

The anthropologist Nick Seaver notes that behaviorist texts, seminars, and thinking are pervasive in the Silicon Valley. He writes facetiously, "Replace the pigeons with people,

build the right levers into your product, and you too might amass a user base of compulsive lever-pressers" (p. 4) [218]. In a recent review of digital behavior change interventions [187], Fogg's Behavior Model was the theory most often cited as the process behind the intended behavior change. Understanding the models in the minds of designers today helps to explain how and why certain technologies come to be.

2.2 — Rational Choice Theory: Users as Rational Actors

Rational choice theory (RCT), developed principally by economists and mathematicians throughout the 20th century, presents a model whereby rational actors maximize their expected utility [169]. When presented with a choice between two options, a rational actor will choose the one that optimizes their "utility," which is often operationalized as wealth or happiness. In RCT, the individual is the elementary unit of analysis--the rest of social life is reducible to the actions of these rational actors [217]. The most useful understanding of RCT is as a baseline model of rational action that predicts some human behavior, particularly economic behavior in the aggregate. In this sense, RCT provides a useful conception of how people act strategically to maximize utility or achieve long-term goals.

RCT is relevant here because it often frames the conversation around sense of agency. Unfortunately, RCT is often overextended as a fully descriptive and/or a normative theory of human behavior. As we shall in the discussion of dual systems theory, it is well-established that RCT has serious flaws as a descriptive theory; a rational actor would never act in 'unintentional' ways that go against their self-interest. Its use as a normative theory for how society should be organized is even more problematic. A belief in humans as highly rational actors, has led many economists, politicians, and technologists to the neoliberal view that the only moral obligation of firms and the government is to provide freedom of choice, i.e., the ability to choose without coercion [72]. What people do with that choice is their own responsibility, but it is assumed that it

will work out for the best because rational actors always make the decision that advances their self-interest.

Like behaviorism, RCT often influences the mental model of designers. In the case of technology design, the common normative overextension of RCT takes the form of conflating what users do with what users really want [218]. That is, because users are rational actors, they must only do things that are in their own self-interest, which in turn validates any action that designers can get a user to take (e.g., spend more time on the site). In reality, the conflation goes beyond just 'what users do is what users really want,' to *'anything we can get users to do is what they really want'*. Only a soon-to-be-unemployed designer, developer, or product manager looks at low time-on-site and concludes 'what users must really want is to not use our product very often, so that's what we should optimize for.'[1] Instead, designers look for ways to give the business model what it 'wants,' which is often getting users to come back more frequently, and then rationalize that as giving users what *they* want.

The incoherence of this position is evident in how companies design for pigeons (i.e., to exploit psychological vulnerabilities), but claim that their users are rational actors (which absolves them of responsibility for any unintentional use that 'happens' to occur). In her rich account of machine gambling in Las Vegas [215], Natasha Dow Schüll describes how gambling developers profess obsession with giving players what they want (usually "entertainment" or "fun"), but tell a different story when they explain what they actually design for: a zone of absorption, in which entertainment disrupts from a steady reward schedule (p. 399). In the case of a social game like FarmVille, the service may exploit social pressure (e.g., Paul and 7 more of your friends want you to play!) or loss aversion (e.g., Check now, your crops are dying!) to get the user to spend their behavior in a way they did not intend and later regret, but justify it as giving the user what they want. Van Couvering describes how equating retention (what users do) with user satisfaction (what users want) helped business people convince developers to go along with various technical changes

[1] There are exceptions, primarily among productivity-oriented tools. Bing's search team optimizes for getting users to click away from their results page as fast as possible [87].

that they were reluctant to make because they were not sure that they were in the user's best interest [237]. Despite its obvious flaws, this self-serving flavor of rational choice theory *('anything we can get users to do is what they really want')* is a powerful framing in the field of product design.

2.3 — Dual Systems Theory: A Model of Cognition

In behaviorism and rational choice theory, the focus of analysis is individual behavior. In dual systems theory, it is the mind. Dual systems theory is a set of theories that emerged from psychological findings of the late-20th century [187]. These theories all identify two distinct systems of cognition, although the exact mechanisms of the cognitive process differ in each.

Dual systems theory provides an empirically-grounded account of self-regulation that can be applied to the digital environment. The psychologist Daniel Kahneman provides a lucid account of the two systems [104]: The automatic system (System 1) is fast, unconscious, and effortless, whereas the reflective system (System 2) is slow, conscious, and effortful (**Table 3**). Automatic system control is driven by environmental inputs and internal states along with cognitive pathways that map the current situation to well-learned habits or instinctive responses [161]. Responses like scratching mosquito bites, or frequent patterns of digital device use like picking up one's smartphone to check for notifications, can happen via automatic system control. Reflective system control is driven by goals, intentions, and rules held in conscious working memory [10,160]. This control is necessary when a goal requires choosing between alternatives or overcoming undesired habits [172], for example if one has a goal of not scratching mosquito bites or not checking a smartphone notification.

Table 3. High-level characteristics of the dual systems

Automatic System (System 1)	Reflective System (System 2)
Fast	Slow
Unconscious	Conscious
Effortless	Effortful

The dual systems model is built in large part on Kahneman and Tversky's work on heuristics and biases, which identified significant and systematic ways in which actual human behavior departs from the rational actor model in which people always choose behaviors that maximize their own self-interest. For example, a rational actor would equally value a loss of $100 as a gain of $100. But in reality, Kahneman and Tversky found that people are *loss averse*: they experience losses more severely than equivalent gain [235], i.e., the pain of losing $100 is often far greater than the joy gained by finding $100.

Another example is present bias: people often choose smaller immediate rewards over larger rewards in the future, a phenomenon that is also known as time discounting. This applies to decisions such as what snack to eat [124,192], how much to save for retirement [92], and which movies to watch [191]. For example, when people choose a movie to watch *this evening* they often choose guilty pleasures like *The Fast and The Furious*. But when they choose a movie to watch *next week* they are more apt to choose a challenging but meaningful film like *Schindler's List* [191]. It follows that one way to understand a loss of control in the digital environment—such as intending to go to sleep, but then watching another YouTube video when it autoplays—is as behavior guided by the automatic system that conflicts with the long-term goals of the reflective system.

2.4 — An Integrative Dual Systems Model

When research interventions in HCI and other fields use dual systems theory, they often fail to realize its full potential. Authors might mention the two high-level systems

and then proceed to choose from a grab bag of heuristics and biases (e.g., loss aversion, present bias) that influence behavior--for example, the popular book *Nudge* itself follows this approach [232]. However, this grab bag approach neglects a deeper understanding of how the automatic and reflective systems interact to shape cognition and behavior.

To provide an understanding of the mechanisms underlying dual systems theory, colleagues and I (in order of authorship: Ulrik Lyngs, myself, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt) reviewed the psychological literature to synthesize an integrative model (**Figure 3**) that shows the cognitive components of reflective (System 1) and automatic (System 2) control [144]. In this model, sensory input (stimuli) from the external world or internal states enter cognition. Automatic filters look out for stimulus properties that are likely to be important either through innate sensitivities (e.g. alarming noises) or learned associations (e.g., a smartphone notification) and boost their signal strength. In this way, some stimuli remain below a threshold of awareness and responded to by the automatic system via reactions or non-conscious habits. However, other stimuli evoke a response strong enough to gain access to conscious attention from the reflective system. The reflective system can then consider relevant goals, calculate the expected value of control (a cost-benefit analysis), and decide upon a course of action (an action schema).

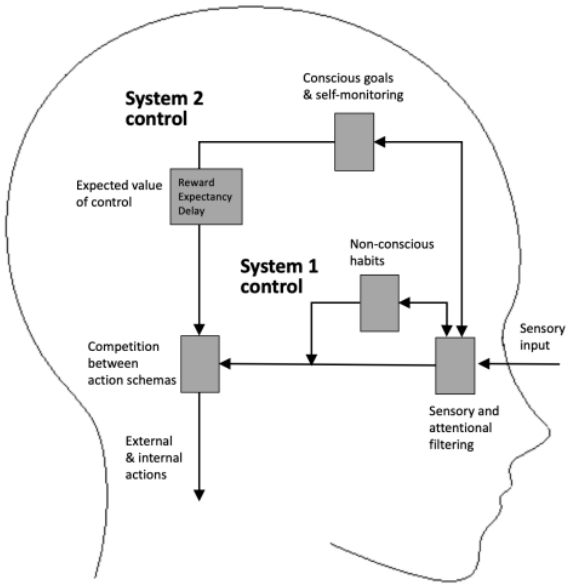


Figure 3. An integrative dual systems model. Automatic system (System 1) control is fast and unconscious, whereas reflective system (System 2) control is slower, conscious, and capacity-limited. The strength of System 2 control is mediated by the expected value of control. Reprinted with permission from [144].

We then applied this model of dual systems theory to 367 digital self-control tools available on stores for browser extensions and mobile apps to identify underexplored opportunities for designers [144]. For example, we found an underexplored possibility of making a distracting site for the user (e.g., Reddit) automatically redirect to a productive one (e.g., Asana), a potential way to scaffold the transfer of reflective system goals to automatic system habits. Another underutilized approach was delay; people are known to be highly sensitive to delays in online contexts [118], so adding a slight time cost could change the user's cost-benefit analysis of visiting a tempting site that conflicts with their long-terms goals. An understanding of the processes underlying the automatic and reflective systems, can help designers identify new opportunities for supporting digital self-regulation.

2.5 — Early Self-Regulation Interventions are More Effective

A popular media narrative that people's inability to control technology use is due to their lack of willpower, a framing of the problem that many users have internalized and that technology companies encourage [123]. After all, rational actors with inexhaustible willpower would just reject any behaviors that did not align with their long-term goals. In this narrative, users themselves are to blame for their failure to resist temptation.

Yet there is an emerging consensus that people who are effective at self-regulation do so primarily by removing temptations from sensory input rather than by resisting through the conscious effort of the reflective system [63,98,144]. In other words, people tend to succeed at self-regulation not via willpower, but rather through advance planning. For example, rather than looking at a cookie that conflicts with dietary goals and resisting it in the heat of the moment, a more effective method of self-regulation would be not to bring cookies into the house so that the temptation never arises in the first place.

If intervening earlier is more effective, what do early interventions look like? Duckworth and Gross propose a model [62,63] that places self-regulation strategies in a sequential order (**Figure 4**):

- *Situational strategies* - changing physical or social circumstances in ways that facilitate control;
- *Attentional strategies* - choosing where to place attention;
- *Appraisal strategies* - changing subjective interpretation of the features that have entered awareness;
- *Response strategies* - voluntarily suppressing undesirable impulses in the heat of the moment (i.e., willpower).

The authors suggest one valuable approach here is implementation intentions, where an individual foresees potential self-control dilemmas and creates an implementation

intention, wherein the next time that temptation comes to attention they already have a tidy plan for how to address it [76]. For example: “When I get the urge to check Facebook during work hours, I will instead read news.” However, the most effective stage at which to intervene would be to modify the situation that users face in the first place. This lesson of intervening earlier rather than later might also be applied to the design of tools for digital wellbeing.

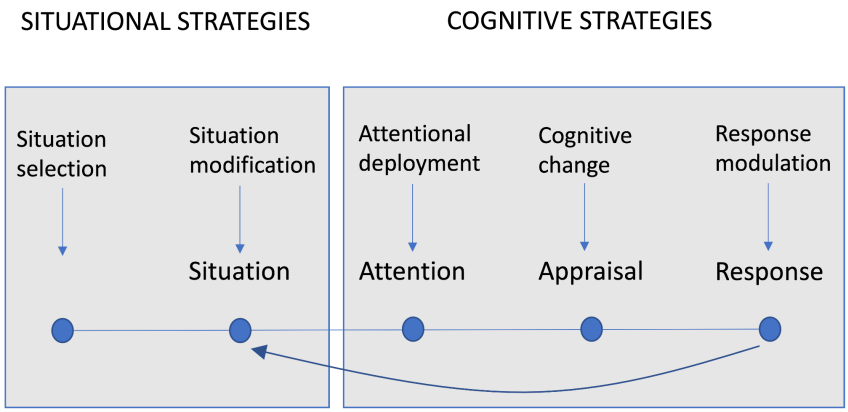


Figure 4. The process model of self-control, adapted from [63]. This model focuses on how ‘impulses’ — response tendencies to think, feel, or act — develop over time, and organizes self-control strategies according to the stage of impulse generation at which they intervene.

CHAPTER 3 — NUDGING, DARK PATTERNS, AND USER FRUSTRATIONS

A first reaction to the known heuristics and biases of the automatic system might be to attempt to shift all cognition to the reflective system, so that behavior is guided by deliberative reasoning. In fact, this is the approach that the vast majority of behavior change interventions in the field of human-computer interaction take. A review of persuasive technologies found that 165 out of 176 research interventions targeted the reflective system, for example by presenting new information to the user to affect their rational decision calculus [2]. By contrast, only 11 out of the 176 research interventions targeted the automatic system. The problem is that access to the reflective system is severely limited by both capacity (e.g., working memory) and availability (e.g., due to fluctuations in emotional state and fatigue), so interventions that target this system alone may not achieve their desired effect [143].

One way to understand the relative strengths of each of the two systems is via the metaphor of the rider and the elephant, introduced by the social psychologist Jonathan Haidt [83]. The rider (the reflective system) sits astride an elephant (the automatic system). The rider can pursue goals, see into the distance, and guide the elephant to move in certain directions. But if the two come into conflict, the rider is at the whim of the much more powerful elephant. The most effective solution is *not* for the rider to train themselves to overpower the elephant in moments of conflict (i.e., strengthen willpower), but rather for the rider to wisely shape the path of the elephant.

This 'shaping of the path' can also be called choice architecture or nudging, an approach that has become highly influential in the design of digital wellbeing interventions. This chapter introduces nudge theory and reviews how it has been applied to digital wellbeing tools. In particular, it considers how nudges are used not only to support digital wellbeing, but also in *dark patterns* that undermine digital wellbeing and how these result in user frustration.

3.1 — Nudge Theory

Where dual systems theory is a descriptive theory of cognition, nudge theory is an interventionist one. Developed by the behavioral economist Richard Thaler and the legal scholar Cass Sunstein, nudge theory applies dual systems theory to the design of choices and environments to achieve behavioral outcomes [232]. They define a nudge as follows:

> *A nudge is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid.*

The premise is that people are *not* rational actors who respond only to incentives, but rather predictably irrational, influenced by known cognitive heuristics and biases. The insight is that rather than relying upon the reflective system—which is slow, conscious, and effortful—designers can use nudges to appeal to the automatic system, which is more readily available and reduces burden upon the user.

In *Nudge*, Thaler and Sunstein call upon choice architects to design for what the user's reflective self would want [232]. Yet how is a designer to know a user's true preferences? This is a vexing problem, both in theory and practice [140]. In the absence of a clear theoretical and practical answer, designers often end up defaulting to *'what we can get users to do is what they really want,'* weakly rationalized by a notion of "freedom of choice" grounded in RCT. This is highly problematic because it often influences people towards actions that are against the interests of their reflective self.

3.2 — Dark Patterns

Although Cass and Sunstein intended nudge theory to be used in the interest of the reflective self, in actuality it is often employed in designs that go against the interests of the reflective self, in which case they might be called *dark patterns*. The design practitioner Harry Brignull [254] introduced the concept of dark patterns to rally the design community against such exploitative designs. Dark patterns can be defined as (a) a

recurrent configuration of elements in digital interfaces that is (b) intentionally created by a designer and (c) leads to user behavior that goes against the user's best interests [133]. This includes interfaces that are designed to deceive the user, and often also those that apply some level of coercion. The designers usually achieves this by exploiting psychological vulnerabilities, i.e., heuristics and biases in user's psychology.

Brignull's original types of dark patterns focus on financial and privacy harms to the user. However, given that people routinely report using technology in ways that are a waste of their time and that they later regret [4,94,114,139], there is a need for research to examine which design patterns prompt such *attentional harms* for the user. We might term these 'attention capture dark patterns', designs that manipulate the user into spending time and attention in an app against their best interests. In my work, I use this framing of *dark patterns* to broaden the appeal of my work to a broader community of design practitioners, researchers, and regulators and extend it to address attentional harms.

Tech industry insiders, like the ex-President of Facebook, warn that social media apps are especially likely to innovate and employ design patterns that "consume as much of your time and conscious attention as possible" [180]. For social games, one such a proposed pattern is "playing by appointment," wherein a player must return to play on a schedule defined by the game, or else lose their precious resources, which exploits the loss aversion bias [247]. For social media, a common suggestion in popular self-help guides is to take back control by turning off notifications [105]. However, it is not yet fully established whether or when such mechanisms lead users to feel a loss of control. For example, some users report that notifications actually reduce their checking habits, since they know they will be alerted when their desired content is ready [179]. My work takes a systematic approach to identifying potential *attention capture dark patterns* within the YouTube interface.

3.3 — User Frustrations

Whether in diet, fitness, or relationships, there is often a gap between how people aspire to act and how they actually act, which is sometimes called the intention-action gap [216]. In the domain of technology use/non-use this gap exists too and is exacerbated by the dark patterns discussed above. As a result, users express frustration and concerns with how they spend their time and attention on technology.

These frustrations can be grouped into specific and general concerns. Specific concerns are ones where users can articulate how their actions connect to undesirable outcomes. General concerns are ones where users express a sense of dissatisfaction, but without pointing to a specific negative consequence.

Specific Concerns

Specific attention-related concerns over technology can be clustered into several areas:

Work: Users often worry that the constant interruption and allure of technology impedes their ability to do work (broadly defined). At the office, employees report checking their email more often than they would like or are obligated to [150]. At the library, students report procrastinating on papers or escaping from challenging problem sets when social media is just a click away [233]. Here designers should be cautioned that people sometimes employ the language of "productivity"--which could be misinterpreted as a specific work-related concern--when their actual concern is about making good use of their time [15,139]; this general concern is addressed below.

Mental and physical health: Losing the ability to focus is a concern; some people relate that they are no longer able to sit down and read a book like they used to [37,170]. This concern is not based on any one instance of use, but rather on the general pattern of multitasking or continuous partial attention that technology encourages [11]. Mindfulness traditions view the ability to focus as the foundation of a virtuous life [55] and see its erosion as a dangerous consequence of technology use [126]. Physical health too is a

concern, as compulsive technology can negatively affect sleep and increase sedentary behavior [121].

Social: Social factors shape this gap in myriad and complex ways. Offline, people report concerns about "phubbing," either ignoring or being ignored by others when they are busy with their smartphones [114]. Online, users report feeling compelled to reply to messages for social or work reasons even when they would rather not [148].

Addiction: In certain cases, the gap between intentions and actions is so wide and detrimental to wellbeing so as to constitute a clinical addiction. In Las Vegas, people attend addiction groups for support by day and then gamble on machines at night while their young children are left to run around a grocery store [215:250]. By comparison, when people speak of smartphone "addiction" in a more colloquial manner, they likely mean an intention-action gap that is far less severe in its consequences for the individual user, but affects a far larger population. Nonetheless, as emerging technologies become increasingly ubiquitous and immersive (e.g., VR), researchers should monitor for signs of clinical addiction.

Designers should consider both subjective experience and expert evidence, and some specific concerns have more support from experts than others. For instance, researchers have found little or no support for health fears about radiation from smartphone signals, but considerable support for worries that nighttime use negatively affects sleep duration and quality [49].

A General Concern: Time Well Spent

My own dissertation research has found that beyond these specific concerns, people often express a general concern about both the quantity and quality of time they spend in apps. Here the concern is not so much that the experience itself was unenjoyable in the moment, but rather that it was a "waste of time" or "unproductive" in hindsight. These experiences are often described in terms of a loss of agency, as in "Getting sucked into some mindless BuzzFeed article" (Lukoff et al., 2018, p. 15). Nor do people necessarily

have a specific alternative activity or goal in mind, but they still look back upon their use with a sense of guilt or regret. In this sense, it is apt that the name of one movement to fight against the attention capture practices of the technology industry is phrased in the past tense: *Time Well Spent*. My dissertation addresses this general concern by exploring how we can design for users to gain a greater sense of agency over their technology use and ensure that their time is well spent.

CHAPTER 4 - DIGITAL WELLBEING TOOLS

Digital wellbeing has become an umbrella term for efforts to support intentional technology use [138]. To identify new design opportunities in this domain, it is helpful to understand what has been tried before in research and industry. I draw upon both my own reading of the academic literature and our review of 367 digital self-control tools on app and web stores from a dual-systems perspective previously mentioned in Chapter 2 [143]. I organize my review according to how the primary design mechanism in each intervention maps to dual systems theory. The first section on **reflective system designs** will discuss tools that employ self-monitoring, goal-setting, or change the expected value of use/non-use. The second section on **automatic system designs** will discuss tools that remove sensory input or block functionality.

4.1 — Reflective System Designs

These designs require the conscious and effortful thinking of the reflective system to be effective.

Self-monitoring. Self-monitoring is common in tools for the intentional non-use of technology. PreventDark, an Android app which detects phone use in the dark and notifies the user that they should put their phone away, explores a context-aware solution for the specific concern of sleep loss [197]. The TimeAware desktop widget presents the same usage statistics in two ways: (1) productive time or (2) distracted time. An experiment found that only the latter frame reduced time spent on 'distracting' apps, but it also increased stress for users [110]. Displaying usage statistics is also popular in commercial tools and is done in both of the major tools that come pre-installed on smartphones: ScreenTime for iOS and Digital Wellbeing for Android.

Conscious goals. Some tools ask the user to set conscious goals and then help monitor progress. MyTime has the user set a daily limit for apps that they would like to reduce their time on. It tells the user when time is up, accompanied by a reminder of their

aspiration for that day [94]. Setting conscious goals is also a component of apps with more coercive enforcement mechanisms, such as GoalKeeper [108].

Social awareness. Social features can also be leveraged in support of intentional technology use. NUGU has users set usage goals and then share performance with friends to receive encouragement [114]. Lock n' LoL lets a group of users synchronously lock their phones, which mutes notifications and restricts usage [113].

Expected value. Other designs alter the expected costs or rewards of technology use or non-use in minor or major ways. This class of interventions has grown popular in HCI research and has been variously termed disruptions, micro frictions, micro boundaries, lockout tasks, and interaction restraints [52,94,181]. If the cost of the disruption is low, these may also function similar to a reminder: for instance, LocknType requires the user to type in a string of numbers in order to access social media apps. If the cost of the disruption is high, the mechanism may be more similar to a shove than a nudge: Goalkeeper sets a daily use limit that is enforced by escalating the amount of time that the user is locked out from their phone (1, 5, 15, 30, or 60 minutes) [108]. TypeOut combines a lockout task with goal setting and affirmation at the moment of temptation [246], finding that these two interventions together significantly outperformed each of interventions on their own in terms of the rate of intervention acceptance and reducing app use. However, goal affirmation did not decrease user frustration with the lockout task, indicating that it is difficult to overcome user reactance against micro frictions.

4.2 — Automatic System Designs

These designs encourage intentional technology non-use without requiring conscious effort on the part of the user. Sometimes these interventions can also break habits of the automatic system, such that the behavior persists even after the intervention is removed [187]. Overall, these approaches appear to be far more common in the design of commercial tools than research ones; few of the academic tools that I reviewed utilized automatic system design mechanisms.

Removing sensory input. One design approach is to minimize or hide undesirable devices, apps, or functionality. Lottridge et al. designed a Firefox extension which reduces the prominence of non-work tabs in the browser, finding that it reduced multi-tasking and increased time on work-related websites [131]. HabitLab, a browser extension that helps users reduce distractions on social media sites, includes some interventions that remove specific elements, such as feed items from sources known for clickbait such as Buzzfeed [117]. In my own work with colleagues, I have designed and tested a browser extension that removes the Facebook newsfeed [145] and built an alternative client for Twitter that hides promoted tweets [248].

However, overall there is still surprisingly little study of how designers might hide or minimize problematic features. This approach is also absent in Hiniker et al.'s design taxonomy of smartphone non-use, perhaps because smartphone apps are 'sandboxed' so as to make this technically difficult. By comparison, on app and web stores, hiding specific features (e.g., the Facebook newsfeed) was present in about one-third of the 367 digital self-control tools (primarily Chrome browser extensions) that we reviewed, suggesting this as a promising opportunity for further research. I explore this approach in my design and deployment of SwitchTube (Chapter 7), an app for watching YouTube videos with a Focus Mode that hides video recommendations.

Blocking functionality. Blocking is perhaps the most popular of all interventions: AppDetox lets users to set times to block certain apps [130], as do about one-third of commercial tools. Interestingly, user reviews often complain that such tools are *not restrictive enough*, and that they contain loopholes that allow the user to escape from their prior commitment [194,233]. In the book *Nudge*, blocking is considered a nudge (or commitment device) if imposed upon oneself, but a shove if imposed by a third party [232].

Additional opportunities. Only a few commercial technologies and none of the research interventions tried replacing undesirable habits with new, desirable ones (e.g., via implementation intentions [76]). None of the tools employed subconscious stimuli (insofar as I am aware) as is done in mindless computing [2].

Beyond Screen Time Tools

Despite some diversity in the design mechanisms employed by digital wellbeing tools, early efforts have primarily taken the form of 'screen time tools' that focus on limiting or blocking time spent on apps. These now come pre-installed on all iOS and Android devices. For example, Apple's Screen Time prompts users to set limits on particular apps (e.g., Facebook) or categories of apps (e.g., Social Media). If time is exceeded, the phone displays a pop-up that prompts the user to leave the app or ignore the limit. Unfortunately this approach also indiscriminately blocks access to features within apps that users still do want to access (e.g., the groups feature in Facebook).

My dissertation work examines the consequences of forcing users into this all-or-nothing bargain and explores an alternative approach wherein one redesigns the features *within* apps that are problematic in the first place (e.g., the Facebook newsfeed). In doing so, I aim to support the user's sense of agency and reduce the regret, guilt, and frustration they feel with the time they spend on digital interfaces.

CHAPTER 5 — UNDERSTANDING HOW APP USE LEADS TO A LOSS OF SENSE OF AGENCY FOR TIME SPENT ON DIGITAL INTERFACES

Technology users often describe how the apps they use are pleasurable, but that they can distract from their higher goals. For example, checking social updates on Facebook might be fun, but take time away from a lunch with friends or a challenging programming assignment [145].

One way to understand the problem is through the lens of Aristotelian Virtue Ethics [100] . Many apps appear to be excellent at providing hedonia, or maximizing what feels good right now. But, the ancient Greek philosopher Aristotle proposed that eudaimonia, or growing through challenging experiences that bring meaning, is a necessary component of well-being. In today's era, how can we create digital environments that support meaningful experiences and digital wellbeing?

To investigate this question, we (in order of authorship: myself, Cissy Yu, Julie Kientz, and Alexis Hiniker) conducted an experience sampling study that asked about the types of mobile app use that participants found meaningful or meaningless that we published at Ubicomp 2018 [139]. Across 45 Android smartphone users and 86,000 samples of app use, we found that productivity and information use cases were, on average, significantly more meaningful for participants than entertainment and social media. However, this effect was moderated by social context: at times, the same type of use was found to be meaningful and sometimes not. One father ordinarily found the game Candy Crush to be meaningless, but in cases when he was overwhelmed by caring for his three children found it meaningful as a momentary form of escape and emotional self-regulation.

Although our initial focus in this work was on meaningful use, our findings suggest that a loss of sense of control over how time is spent is central to meaningless app use. In this dissertation chapter, I therefore reconsider our results not just from the perspective

of virtue ethics, but also in terms of sense of agency. I address the question: What app use makes people feel a loss of agency for the time they spend on digital interfaces? **(TQ1)**

5.1 — Background and Motivation

Meaningful and Meaningless Technology Use

Supporting smartphone users who express dissatisfaction requires a nuanced understanding of when and why users are dissatisfied. Understanding the usage patterns that leave people frustrated with their own choices has the potential to help users shape the strategies they use to manage their usage behaviors. It also has the potential to help designers avoid design patterns that satisfy short-term metrics (e.g., time spent in the app) at the expense of long-term declines in user satisfaction.

To understand user dissatisfaction, one common approach is to examine which types of media use leave people 'feeling bad'. In studies of media use, this negative feeling is often measured in terms of emotional valence [12,119,238]. This approach aligns with a hedonistic tradition that views subjective well-being (and sometimes happiness too) in terms of affect balance: the presence of positive affect and absence of negative affect [251]. Yet affect balance alone may not fully capture why people are dissatisfied with aspects of their smartphone use [140].

In this study, we focus on the construct of meaningfulness, a key component of the eudaimonic conception of well-being. It is true that meaningfulness is positively related to affect balance: most meaningful experiences also make us feel good. However, the two concepts still differ in important ways [12,249]. For example, scrolling through pictures of adorable kittens might cheer one up, but feel meaningless. Conversely, messaging one's ex-girlfriend to apologize might be a sad experience, but meaningful. In contrast to hedonic happiness, people associate meaningfulness with giving rather than taking in relationships, enduring unpleasant experiences in pursuit of future goals, and reflecting one's ideal self [12].

The field of positive psychology made great progress in understanding how to cultivate happiness in the hedonic sense, yet it is only more recently that scholarship has turned to the eudaimonic notion of meaningfulness [12]. Similarly, in positive computing [34], valuable work has investigated happiness, often drawing upon the sensing technologies advanced by affective computing [33], yet there seems to be less emphasis on studying meaningfulness. This paper focuses on the construct of meaningfulness as a key consideration when designing for human flourishing.

Using a Uses and Gratifications (U&G) perspective [201], we explored the meaning, or lack thereof, that people derive from their smartphones through interviews with users and log data capturing their experiences in the moment. We built a mobile app that logged app use and used the experience sampling method (ESM) [54] to capture the behavior of 45 smartphone users for two weeks. We asked smartphone users about their underlying motivation for phone use at the start of, during, and at the end of specific instances of app use. Using qualitative data from these ESM prompts and from exit interviews, we extracted themes about the gratifications users seek as they engage with their phones and the meaning they draw from these experiences. We then corroborated users' subjective descriptions of their experiences through a quantitative analysis of 86,402 sessions of app use and the 9,318 sessions from which we collected experience samples.

Participants reported that habitual use was typically less meaningful than intentional use. Habit-driven experiences made users feel a loss of sense of agency over their own behavior. Quantitative analysis of log data was consistent with these themes, showing both that users turn to their phones out of habit and lose track of their intentions and fall into habitual patterns of engagement with their phone. Participants further said that they turn to their phones to escape from negative emotions and held mixed opinions about how well such engagement served them. By examining habit, type of use, and meaning together, we provide insights into why people engage in behaviors they wish they could change.

The Uses and Gratifications Perspective

Historically, new technologies have been examined from a media effects perspective, but beginning in the 1950s, communications researchers began to examine people's motivations for engaging with technology, rather than simply examining the effects of technology and treating users as passive recipients of media experiences [202]. The resulting Uses and Gratifications (U&G) theory explains that media use is an active choice on the part of a user, driven by the user's desire to seek specific gratifications.

U&G has proven to be a productive frame for understanding users' experiences with a wide variety of new technologies. Prior work has drawn on U&G to explain why users engage with VCRs ([46], social media ([189]), early cellular phones [125]), soap operas [201]), tablets ([146]), and many other technologies. While other work has looked at the gratifications users obtain from their smartphone use (e.g., [96]), this investigation uses them as a lens for examining technology resistance and users' frustration with their own behavior.

At a high level, the gratifications users seek from media can be grouped under two motivations: instrumental motivation, in which the user engages with technology intentionally to achieve a specific purpose, and habitual motivation, in which the user engages with technology habitually to pass the time ([199,200]. This dichotomy is particularly relevant to this investigation because prior work suggests that habitual motivation is associated with the types of smartphone use that people would like to reduce. For example, SNSs are used habitually ([196,208]) and are common sources of lagging resistance for users [13,96]. Other work shows that apps with high informational rewards lead to checking habits [179]. Thus, it is possible that the instrumental-habitual divide contributes to the extent to which users draw meaning from their experiences.

We can also consider the gratifications that users seek from a particular technology at a more detailed level that may be more actionable for designers of mobile apps. Research studies that collect large volumes of data about smartphone use tend to rely upon the categories provided by commercial app stores like Apple's App Store and

Google's Play Store [74,245], sometimes making manual adjustments ([21,190]. However, app store categories are only a crude approximation of the gratifications that smartphone users seek. First, categories can be ambiguous. For instance, the top rankings in the "Lifestyle" category on the Play Store include apps for dating (e.g., Tinder), retail (e.g., Starbucks), tools (e.g., Timely Alarm Clock), and religion (e.g., Salatuk - prayer time). Second, any app-level categorization scheme will not capture how a single app can be used in multiple ways that do not fall under a single category. For instance, as of August 2017, the top-ranked app in the "Communication" category on the U.S. Google Play Store was a mobile browser (Chrome), which can be used for communication (e.g., emailing a friend), but also for information-seeking (e.g., searching on Google) and entertainment (e.g., watching videos on YouTube). A more user-centered typology is needed to understand what people want when launching an app.

Instead of app store categories, typologies based on surveys and experiments with users [20,30,40,115,238], better inform the U&G perspective on smartphone use. Bessiere et al. used surveys and factor analysis to identify four distinct ways of using the internet: communication with close ties, communication to meet people, information-seeking, and entertainment ([20]). Chan developed a similar typology for smartphone use, also with 4 types: voice communication, online communication, information-seeking, and passing time ([40]). He found that communicative uses were positively related to subjective well-being, whereas non-communicative uses were not. Verduyn et al. compared social media use that was active (direct exchanges) and passive (consuming information), finding that passive Facebook use led to declines in affective well-being [238]. The types used in this prior work informed the development of our own U&G typology of smartphone use. In particular, their findings suggest that active communication and passive consumption of social media should be seen as distinct types of use that offer different gratifications.

We build on prior work by using a U&G perspective to examine the meaning that people derive from smartphone use. We study what smartphone users want at two levels:

at a high level, in terms of instrumental and habitual motivations, and, at a more detailed level, in terms of five different types of use (see Methods).

Research Questions

Our research questions were as follows:

- **RQ1:** Which U&Gs are associated with smartphone use that people find meaningless?
- **RQ2:** How do U&Gs change as a user engages with an app?

5.2 — Methods

We used mixed methods to collect both smartphone log data and multiple sources of qualitative data about users' experiences. 45 adults living in the United States installed an Android app that logged their smartphone use and asked them questions about their experience. All participants completed an exit survey, and 11 also participated in an exit interview.

Recruitment

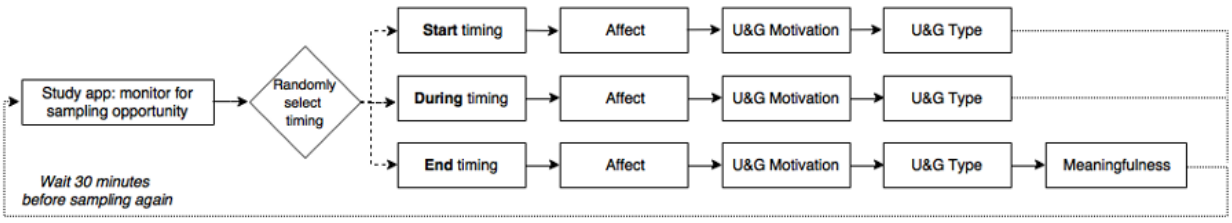
Recruitment In June 2017, we recruited 45 U.S. adults from Craigslist (n=34) and university email lists (n=11). We did not solicit users who had particular desires to reduce smartphone use, but rather wanted to recruit a general U.S. population. Participants responded to an ad for a study that "tracks the apps you use and asks you about your experience." Median age was 28 (range 18-50). 53% identified as men and 47% as women. Racial identity (non-mutually exclusive) was 77% White, 13% Asian, 4% Black, and 9% other, with 16% of all participants indicating Hispanic/Latino ethnicity. In terms of education, 13% held a high school degree or equivalent, 36% had completed some college or an associate degree, 36% held a bachelor's degree, and 16% had a master's degree or doctorate. Prior to this general recruitment, we had also conducted two iterations of pilot testing (n=4, n=5) to ensure that study instructions were clear and to troubleshoot our study app.

Participants were all Android smartphone owners (version 5.0 or later) who reported using their phone for at least 30 minutes per day (median=165 minutes). Our study app was made available on the Google Play Store and required participants to enter an invite code upon installation. Two invited respondents were unable to install the study app due to technical issues and could not continue as participants. Participants were compensated with a $75 Amazon gift card and received an additional $25 if they completed the exit interview. This research was reviewed and approved by The University of Washington's Institutional Review Board.

Experience Sampling Method (ESM)

We developed a study app that logged the name, category, date, time, and duration of each app that a participant used during the study period using the android.app.usage API [86]. The code for the study app is available on Github[2]. When the study app detected that participants were using their phone, it asked them short self-report questions as per ESM. Our prompt could be dismissed by pressing the back button; study instructions informed participants that skipping a prompt in this manner was okay if their phone use was time-sensitive. Our sampling protocol is shown in **Figure 5** and described in the text that follows.

Figure 5. Sampling protocol for the study app



[2] https://github.com/yucissy/Emotion-sampling

To capture a diverse set of apps, if an app was sampled, it was not sampled again until after 3 other apps had been sampled. We were concerned that some messaging apps are checked so frequently (e.g., SnapChat) that they would crowd out samples from all other apps [21,206].

The study app prompted participants at one of three random **timings**: *start*, *during*, *and end*. Each instance of app use was sampled at a single timing, not at all three, as we expected that the disruption of our prompt would distort subsequent behavior so that any additional measurement of the same use would be invalid. Although we never collected more than one sample for any particular instance of app use, collecting a large number of samples would allow us to stitch together a picture of how user motivations evolve from the start-to-middle-to-end of app use.

Start samples came at the launch of an app. *During* samples were requested at a random time between 15-120 seconds of use. If a participant closed an app before reaching that number of seconds, the study app went back to the beginning of the sampling protocol. We therefore collected fewer samples during app use than at the other timings. *End* samples were collected immediately after an app's window was closed, provided that the app was used for at least 15 seconds. We selected 15 seconds as the minimum cutoff, as wanted participants to interact with an app before we questioned them about their experience. We took guidance from prior literature that found 15 seconds to be the amount of time associated with at least brief interaction with apps, whereas less time was associated with simply glancing at a device ([9]). Again, if an end sample was selected but app use did not reach 15 seconds, the study app reverted to the beginning of the sampling protocol. This also explains why we collected more samples at the start timing than at the end timing. So as not to overburden participants, we waited for at least 30 minutes after sending a prompt (regardless of whether or not it was answered) before sending another one.

We also measured affect as has been done in prior studies of social media use [119], [238]. Although it was not the focus of our study, we thought it might yield

interesting comparisons against prior work. Affective state was measured along the two dimensions of the circumplex model of affect ([203]): valence (7-point scale; negative to positive) and arousal (7-point scale; low energy to high energy).

U&G motivation asked whether the participant was using the app instrumentally *("To achieve a specific goal")* or habitually *("To browse, explore, or pass the time")* as done in ([96]). *"Not sure"* was also an option.

Our U&G types (see Table 1) were based on prior literature and our own testing. Similar to Bessiere et al. ([20]) and Chan [40], we included communication, information, and entertainment as types of use. We adopted the distinction between active and passive social media use, reusing language from the study instructions in [78]. In our typology, however, active social media use was subsumed under communication. Our aim was for participants to include all communicative use under communication, even if it occurred in apps that are not always regarded as "social media." For example, chatting with friends in mobile messaging apps (e.g., WhatsApp). Based on usage behavior in prior studies [21,245], we added productivity as an additional type. "Not sure" was also included as an option. Although this paper refers to types of use by their short name (e.g., communication), participants were always shown the full name in the study app (e.g., communicating or interacting with other people), as listed in **Table 4**.

Table 4. U&G types with examples

Type of use		Examples
Short name[3]	**Full name**	
Productivity	Getting things done or self-improvement	· Checking balance in a banking app · Logging a run in a MyFitnessPal
Information	Getting information	· Searching on Google · Checking the forecast in a weather app
Communication	Communicating or interacting *with* other people	· Liking or commenting on Facebook · Messaging with friends on SnapChat
Entertainment	Entertainment	· Watching videos on YouTube · Playing games like Candy Crush
Social media	Browsing social media *without* interacting with other people	· Scrolling through the newsfeed on Facebook · Reading friends' stories on SnapChat
Not sure	Not sure	Anything that doesn't easily fit into one of the other categories

To assess our typology, we ran one test of validity and another of reliability. Key methods and results are presented here; for a detailed explanation please see the supporting materials that are published together with the archival version of the published paper [139]. To test validity, we recruited 64 smartphone owners from Amazon Mechanical Turk. We asked participants to describe 3 recent app uses and categorize them according to our typology. In 5 instances, participants used "Not sure" to classify app uses similar to "recording my gym sessions and bike rides" and "I used the application called Memrise to self-teach some Japanese." We therefore broadened productivity from "Getting things done" to "Getting things done or self-improvement." We chose not to create additional types for a few other uses that participants reported difficulty categorizing, such as online shopping (2 cases), flashlight (2 cases), and taking photos (2 cases), as these were less common. We also created examples that we shared with participants to help them distinguish between types of use (see Table 1).

[3] Short names are used in this paper, but participants always saw full names. Participants saw these examples in their training.

To test reliability, we recruited 8 new participants from Amazon Mechanical Turk. Thirty cases of app use were randomly selected from those generated by participants in the test of validity. Each participant completed a 5-minute training for our revised typology (with examples) and then categorized the 30 cases. Light's Kappa, a measure of interrater reliability for fully crossed designs with more than two raters ([84]), was 0.74 (where 0.6-0.8 can be interpreted as indicating substantial agreement ([127])). The same online training was later administered to participants who enrolled in the main study.

Meaningfulness was asked as, ‘‘How much do you feel like you have spent your time on something meaningful?’’ (7-point scale; not at all meaningful – neutral - very meaningful). We did not further define meaningfulness, as we wanted participants to interpret it as relevant to their own lives, as has been done in previous research in positive psychology [12]. The same question was asked about Facebook use in ([205]), in which answers were found to be highly similar to answers in response to alternate wordings “...wasted your time?” and “...done something useful?” (Cronbach’s alpha = .91). This question was asked only at the end timing, as we wanted to allow participants to complete their app use before reflecting upon its meaningfulness.

Exit Survey

Participants completed a short exit survey that asked how often they wished to use their phone for each of our types of use (5-point scale; far less often - the same amount - far more often). Our aim was to compare global intention for smartphone use against in-the-moment ESM reports of meaningfulness. We surveyed global intention at the end of the study rather than the beginning because we did not want this question to influence ESM reports. However, it should be noted that there may have been an effect in the opposite direction: after the two-weeks of ESM prompts, participants may have gained greater awareness and formed new intentions for phone use. The survey also inquired whether participants had experienced technical difficulties using the study app.

Exit Interviews

We chose 12 study participants (6 men, 6 women; age range 18-46) to invite to the exit interview using multistage sampling. In the first stage, we clustered participants into two groups based on gender identity. In the second stage, we further clustered the two groups into sextiles based on age. We then randomly selected one participant from each of these 12 groups. One participant was unable to participate due to schedule conflicts; all other invited participants chose to participate in the interview. We conducted interviews on the phone, and each lasted about 45 minutes.

In the first part of the interview, we asked participants what they had learned about their smartphone use, why they found certain types of use to be more or less meaningful, and whether they wanted to make any changes to their phone-use habits. In the second part of the interview, we made use of a retrospective interviewing technique ([159]). For each participant, we prepared a timeline of their previous 24 hours of smartphone use, and we reviewed it together with them to contextualize specific instances of phone use and to better understand how different types of smartphone use fit into their daily life.

Data Analysis

Interviews. We began our data analysis with the qualitative data from interview recordings that were professionally transcribed. The first and last author first divided interview transcripts and each conducted an open-coding of an independent subset. We then met to collaboratively discuss emergent themes. Using these emergent themes, researchers then repeatedly reviewed transcripts to refine themes and pull out examples. Using example quotes, we conducted collaborative affinity diagramming, clustering examples into a collection of cross-cutting themes. We used these to define our quantitative analysis of ESM and log data by examining whether this log data supported or refuted the themes generated from our qualitative analysis.

ESM Data. We then proceeded with rough confirmatory analysis using ESM data. Rough confirmatory data analysis lies between exploratory data analysis and

confirmatory data analysis, using probabilistic approaches such as confidence intervals and significance tests to conduct an initial assessment of plausible models ([16]). We present most of our quantitative results using simple graphical representations. When we conducted statistical tests, we were interested in variables that were highly significant ($p < .01$) and had large effect sizes, and thus should therefore clearly be included in the model. Our quantitative results should be interpreted as an initial assessment of multiple potential models rather than as confirmation of a single model.

5.3 — Results

We first present descriptive statistics for the ESM data to characterize the smartphone use of participants. Participants used apps a total of 86,402 times over the two weeks of the study. We sampled the experience of participants in 9,318 (10.8%) of app uses. The completed response rate was 86% and the median time taken to answer a sample was 11 seconds. Participants received a mean of 13.3 prompts per day (sd=6.3). Because we sampled only when a participant was using an app, more frequent smartphone users received more prompts.

Table 5 shows a summary of these ESM samples. The majority of use was instrumental (60.5%) rather than habitual (39.5%). Communication was the most common type of use (31.3%) and we excluded “Not Sure” responses from further analysis.

Table 5. Summary of ESM samples of smartphone use

	Samples	Adjusted Percentage
Total app uses	86402	100%
Non-ESM	77084	89.2%
ESM	9318	10.8%
Timing		
Start	4790	51.4%
During	1578	16.9%
End	2950	31.7%
U&G motivation		
Instrumental	4573	60.5%
Habitual	2984	39.5%
Not sure	493	NA
U&G type		
Productivity	930	12.4%
Information	1702	22.7%
Communication	2346	31.3%
Entertainment	1401	18.7%
Social media	1106	14.8%
Not sure	551	NA

The summary statistics for our affect and meaningfulness measures are shown in **Table 6**. There was a positive bias in emotional valence ratings, which has also been found in prior studies ([91]).

Table 6. Summary of affect and meaningfulness

	Number	Mean	SD
Affect			
Valence	8149	3.4	1.2
Arousal	8149	3.0	1.2
Meaningfulness	2230	3.0	1.4

We also examined whether participants reported using the same app in different ways, which was our key motivation for developing a typology of smartphone uses and gratifications. **Table 7** shows the results for the 5 apps for which users answered the most ESM prompts. For each app, we list the total number of samples and the percentage share by type of use, with the largest share in bold. For example, participants used Chrome for

information-seeking in 50% of samples, with the rest of use spread across entertainment (20%), social media (11%), communication (10%), and productivity (9%). Among all 66 apps in our dataset with 20 or more samples, we found that the median of the largest share for a type of use (e.g., information in the Chrome example) was 58%.

These results clearly show that people often use the same app for different types of uses & gratifications. This would not be evident based on app store categories alone: for example, Gmail is listed in Communication on Google Play, but in 39% of samples participants responded that their use was best described as getting information. Snapchat is listed under Social, but in 66% of samples participants reported that they were engaged in active communication rather than passive social media use. For researchers who are interested in the different ways that apps are used or are considering adopting a similar methodology, we anonymized and aggregated this part of the dataset and made it available in the supporting materials of the published paper [139].

Table 7. Top 5 most popular apps and the diverse ways in which they are used

App name	Category in Google Play	Total samples	Productivity	Information	Communication	Entertainment	Social media
Chrome	Communication	701	9%	**50%**	10%	20%	11%
Facebook	Social	476	1%	5%	25%	14%	**54%**
Gmail	Communication	459	12%	**39%**	31%	10%	8%
Instagram	Social	360	2%	5%	26%	21%	**46%**
Snapchat	Social	326	1%	2%	**66%**	11%	19%

In the exit survey, study participants shared two notable concerns about our experience sampling protocol. Three participants expressed concern at receiving prompts while they were driving. In the future, researchers who sample app use on smartphones should consider disabling prompts when people are using navigation apps. One participant expressed frustration at prompts that occurred during SnapChat use, when he had limited time to view a friend's photo before it disappeared. In hindsight, we would

have more strongly emphasized to participants that they were free to dismiss prompts that came at inopportune times.

Meaningfulness of Different Types of Use

Participants held similar opinions about which U&G types were more and less meaningful. Productivity and communication with close ties were mostly considered meaningful, while social media and entertainment were frequently viewed as meaningless. In the exit survey, participants expressed the desire to reduce the use of the same U&G types seen as meaningless.

—Interview Data—

Meaningful Experiences with Smartphones. Participants associated using their smartphone for productivity with a sense of meaning. Productivity use cases were often related to office work: *"The Slack app, that one is meaningful because that one actually gives us support… it's meaningful in that way because it helps me get work done easier"* (P10). For P10, the two weeks of participation helped him realize how he used his phone more productively than he had thought: *"I always knew that my phone is probably more important than my wallet at this point in my life, but I didn't realize how often I used it productively versus just screwing around"* (P10).

Because our definition of productivity was broad ("Getting things done or self-improvement"), participants also reported productivity uses beyond the workplace:

> *A fitness app, that would be something productive, or something that had to do with organization, like creating a to-do list, or, you know, keeping up with plans at work and writing down notes and projects. I'd say that feels a lot more meaningful to me than just maybe going into an app and scrolling through cute cat pictures. (P7)*

It was common for participants to contrast productivity against other use cases such as social media: *"I have a Fitbit and then a calorie counter thing. Using both together I think is*

more meaningful and worth my while than just looking at Facebook" (P6). Asking participants what they found meaningful helped reveal what they found meaningless.

Active communication was also frequently regarded as meaningful regardless of which app store category it fell under:

> *Probably the most meaningful task that I used my phone for was communicating or interacting with other people. And that was irrespective of what kind of app I used, so if it was like Messenger or Facebook or Snapchat, it didn't matter what app I was using, or email. If I was using it to interact with someone, I think I associated that usage of the phone more positively. (P3)*

Of the apps mentioned in this quote, only Messenger is listed in the "Communication" category on Play Store, whereas email apps are usually in either "Communication" or "Productivity," and Facebook and Snapchat are in "Social." Other participants reported meaningful communication in apps that fall under categories such as "Tools" (e.g., the dialer), "Music and Audio" (e.g., Spotify), and "Lifestyle" (e.g., Tinder). Meaningful communication was not bounded by app store categories.

Participants saw interactions with close ties as especially meaningful:

> *I didn't speak to my daughter for at least four or five days. I got my own apartment, I just moved into this apartment where I'm at now, and she's mad because I left. Then yesterday she finally talked to me on Facebook, so that's why I put a 6 [out of 7 for meaningfulness]. (P10)*

When we asked participants what type of use they considered meaningful, it sometimes took them a few moments to respond. Communication with family was often the answer these participants gave upon reflection:

> *Meaningful... I mean other than maybe just calls with family or texting throughout the day, but yeah I would say just generally... Yeah I'd say calls are probably the only really very meaningful thing that I'd be doing, maybe emails with family news or updates like that. (P11)*

We did not distinguish between communication with close and weak ties in our U&G types, but it appears to relate to the meaningfulness of smartphone use.

Meaningless Experiences with Smartphones. Almost all participants singled out passive social media use when asked if they found any smartphone use meaningless:

> *Oh yeah. Browsing social media. Going to Facebook just to browse and not doing anything else. Going to Instagram just to browse. I think those are pretty much meaningless. (P3)*

As in this quote, qualifier words like "just" and "kind of" were used almost every time that participants described their social media use, in a way that downplayed the significance of their activity. This contrasted with the more direct language used to describe more meaningful uses that was free of such qualifiers.

Several participants suggested that social media could be more meaningful if they engaged more actively:

> *Social media has the potential to be meaningful, but the way that I often use it is meaningless, as a way to just pass the time. I am not very active on social media, engaging with it. I am just a consumer which feels less useful to me… not building meaningful connections. (P1)*

P6 was similarly critical of the passive nature of his social media use, using words like "not do anything" and "sitting there":

> *I would just scroll through Facebook and not do anything. I was just kind of wasting time, sitting there, looking at pictures and videos. Looking back on it, no, that really wasn't meaningful whatsoever... Maybe if I was actually going to be using Facebook or texting more or Snapchatting more, it would be actually to interact with someone and speak with someone, so I have better relationships with people, instead of just looking at pictures and not doing anything, interaction-wise. (P6)*

Similarly, in P7's mind, what she called her "lurking" did not build relationships in the same way as "actual interaction" (active communication in our types of use). P3 also regarded his present social media use as mostly meaningless, but believed that it had potential to be more active and meaningful.

Entertainment was also sometimes cited as meaningless. "I wouldn't say any time spent on that app [the Deep Town game] was meaningful, it's just a way to entertain

myself mindlessly" (P11). Another participant explained why he considered entertainment uses to be meaningless: "I just tended to gain more out of using it as a tool rather than using it to pass the time or using it to do something that didn't have like an end gain" (P3). As in this quote, participants often attributed meaningfulness to a lack of 'productive' output.

Experience Sampling Data. Next we investigated whether type of use affects meaningfulness in the experience sampling data. In addition, based on prior literature, we examined how U&G motivation (instrumental versus habitual) relates to meaningfulness.

Figure 6 reveals a clear relationship between meaningfulness and type of use. Productivity, information, and communication show higher ratings of meaningfulness than entertainment and social media. These results are consistent with the interview data. **Figure 7** shows that meaningfulness is also strongly related to U&G motivation. Habitual use is associated with less meaningful experience.

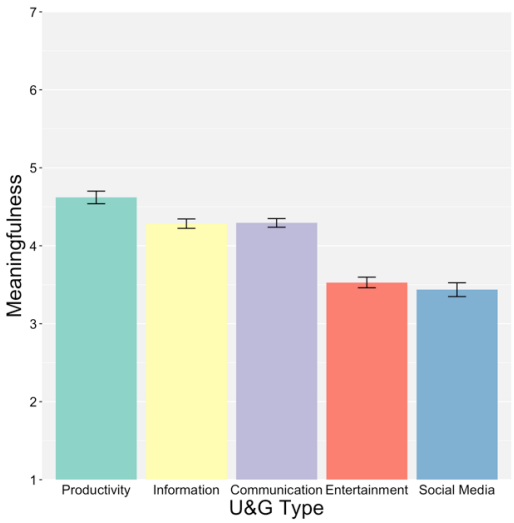


Figure 6. Meaningfulness vs. U&G type. Shows the mean and its 95% CI. Entertainment and social media are related to a lower sense of meaningfulness.

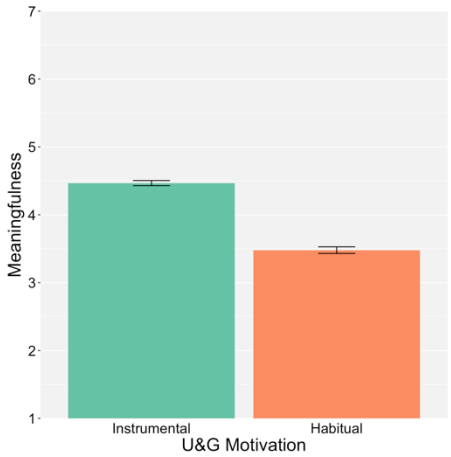


Figure 7. Meaningfulness vs. U&G motivation. Shows the mean and its 95% CI. Habitual motivation is related to a lower sense of meaningfulness.

We performed a linear mixed effects analysis of the relationship between meaningfulness and our explanatory variables (**Table 8**). As fixed effects, we entered U&G type, U&G motivation, and time of day. Time of day was included because prior work suggests that people use their phone less purposefully later in the day [96]. As a random effect, we included participant ID, so that intercept was allowed to vary by participant. We used R and the lmerTest package ([120]) to run models with the REML estimation and test for significance.

Relative to smartphone use for productivity, social media use decreased ratings of meaningfulness by about 0.67 ± .12 (standard errors), holding other factors constant. Entertainment use led to a roughly similar decline (0.56 ± .11) on the 7-point scale for meaningfulness, whereas information and communication were not significantly different from productivity use. Using one's phone habitually rather than instrumentally reduced meaningfulness ratings by about 0.68 ± .07. Coefficients for time of day were mostly in the expected direction (with evening and night corresponding with less meaningful use), but effect sizes were small and insignificant.

Table 8. Mixed Effects Regression for Meaningfulness (n=1947)

Source	*Estimate*	*SE*	*t*	*P*
Intercept	3.74	0.13	35.28	<.001
U&G Type (baseline = Productivity)				
Information	-.13	.10	-1.32	<.188
Communication	-.09	.09	-.95	<.345
Entertainment	-.56	.11	-5.06	<.001
Social media	-.67	.12	-5.72	<.001
U&G Motivation (baseline = Instrumental)				
Habitual	-.68	.07	-9.47	<.001
Time of day (baseline = Morning (6am-12pm))				
Afternoon (12-6pm)	.08	.07	1.23	.218
Evening (6pm-12am)	-.06	.07	-.91	.365
Night (12-6am)	-.17	.11	-1.55	.121

Exit Survey. In the exit survey, participants rated how often they wanted to use their phone for different U&G types on a 5-point scale, ranging from far less often to far more often (Fig. 4). An analysis of variance (ANOVA) on these ratings yielded significant variation among conditions ($F(4, 225) = 26.41$, $p < 0.001$). The results of a post hoc Tukey test are also shown in **Figure 8**.

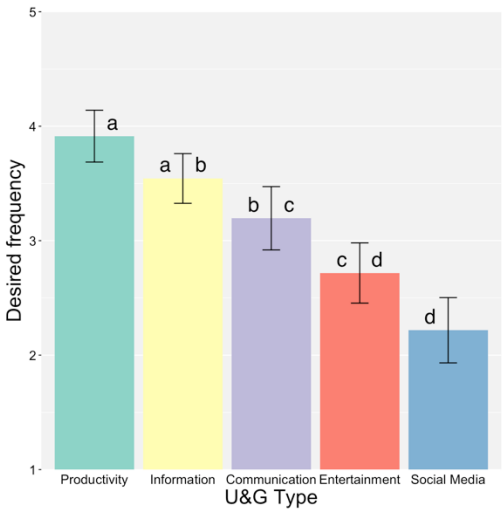


Figure 8. Desired frequency of use vs. U&G type (1 = far less often, 5 = far more often). Shows the mean and its 95% CI. Letters (a, b, c, d) indicate pairs that are *not* significantly different from each other. For example, Productivity and Information are both marked (a), so do not differ significantly. However, Productivity does differ significantly from all other U&G types ($p < 0.01$).

Ratings were mostly consistent with the ESM data, with productivity receiving the highest average rating and social media the lowest. One exception is that communication lay in the middle, with no significant difference relative to entertainment ($p = 0.07$) and informational use ($p = 0.32$). Prior literature and our interview data suggest that communication with close ties would have been rated higher than interaction with weak ties, but our definition of communication as a type of use did not make this distinction. On the whole, participants wanted to reduce usage behavior that was also associated with a lack of meaningfulness in the ESM data.

Loss of Sense of Agency

In interviews, participants reported a loss of sense of agency for the time they spent using smartphone apps. They highlighted the automatic nature of their checking habits, particularly for social media, entertainment, and sometimes communication. This kind of time spent on habitual smartphone use was frequently characterized as meaningless.

Experience sampling data reveal social media and entertainment are used habitually far more often than other types of use. These data also show that participants began smartphone use with a higher sense of instrumental motivation than they ended with. In other words, intention is eroded during the course of smartphone use.

Interview Data

A recurrent theme was loss of sense of agency, acting without the experience of choice [58]. Many participants felt like they were not in control of their use: “A lot of the times I'd be on my browser and maybe I can get sucked into some mindless *BuzzFeed* article. I feel like it's kind of more of a black hole than if you go on a fitness app that's strictly for one thing” (P7). Another participant wished that he had more self-control, particularly when his wife was annoyed with him for being on his phone all the time. However, when asked why he thought he lacked this self-control, he clarified that it felt like his use was no longer an active choice:

> *I'd say it's more just a condition thing at this point. It's so normalized to have the phone in my pocket or on the table or whatever and just be absentmindedly glancing at it or picking up and scrolling through something. That probably isn't meaningful at all, just kind of a force of habit. (P11)*

Lack of sense of control was rarely attributed to active failure to resist in-the-moment, but rather to unconscious habit. We therefore use loss of sense of agency rather than loss of self-control to describe this feeling of acting without the experience of choice.

Participants shared many anecdotes in which they were not even aware of their own use at the moment: “Without even realizing it, I pulled out my phone and just started mindlessly checking my email. Then when I put it away I realized, ‘Oh, I feel bad for ignoring my friends’” (P1). The presence or intervention of close ties often helped a participant become aware of use that their reflective self also considered problematic. One participant told her fiancé that she was not going to check Facebook on her phone while they were watching television together on the couch:

> *I would not look at my phone for about 10 minutes, and then I wouldn't even realize it, but I had it back in my hand and I was looking at it, and then [my fiancé] would bring up, 'You're using your phone again. You said you weren't going to use your phone when we were watching this episode.' I'm like, 'Oh, yeah, right. I forgot,' then put my phone back down. It's gotten better maybe, but it's something that's just a habit. It's really hard to break habits. (P6)*

Smartphone use was viewed as a strong habit and sometimes as an 'addiction'. One participant explained why he wanted to use his phone less often after participating in the study:

> *Because I kind of realized how glued I was. It's almost second instinct, you know? The minute you are bored or don't have anything to do, you just turn to your phone, and I'd like to be able to get out more and really live life. (P7)*

Despite their best intentions, participants reported that because phone use habits had become so ingrained, they were now "really hard" to change.

A lack of sense of agency went hand-in-hand with a lack of meaning for participants. Again, P6 shared: "Sometimes, I'll look at the videos, yeah, and they make me sad, so why would I even, why do I keep looking at it? I don't know why I do, but I do." One participant summed up this relationship, "Candy Crush is absolutely addictive, and it's just absolutely useless" (P5). Other participants described cases of unconscious use as "completely pointless" and "a waste of time."

Experience Sampling Data

To check whether participants were "sucked into" meaningless use, we tested the relationship between U&G motivation and the timing of the sample (**Figure 9**). A chi-square test of independence revealed that the relation between these variables was significant, $X^2(2, 7557) = 109.74$, $p < .001$. The percentage share of instrumental motivation declined from the start timing (65.6%) to the during timing (51.4%). There was a slight increase in instrumental use from the during timing to the end timing (55.7%), which may be because participants may have looked back upon an app use that started with instrumental motivation and later changed to habitual motivation and concluded

that, on the whole, it was still more instrumental. At the start of app use, participants had a specific purpose in mind more often than during or at the end of use.

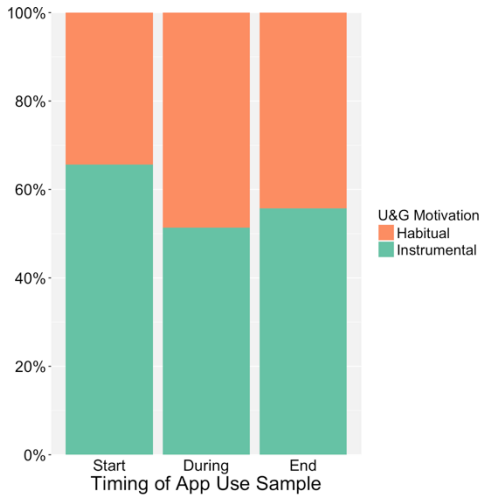


Figure 9. U&G motivation vs. timing. Instrumental motivation declines over the course of use, relative to habitual use.

We also examined whether the erosion of instrumental motivation differed between types of use. **Figure 10** shows the percent of instrumental use and the sample of the timing, for each type of use. First, we note that different types of use had different baselines. Regardless of timing, productivity, information, and communication use was far more likely to be instrumentally motivated than was entertainment and social media.

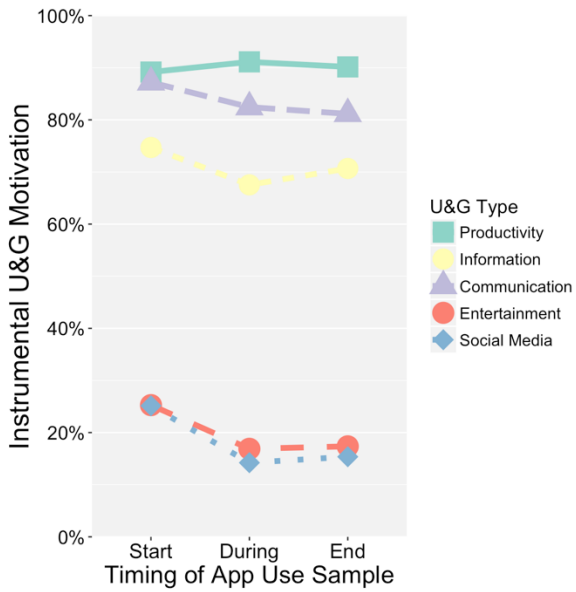


Figure 10. Instrumental motivation vs. timing, by type of use. Compared to other types of use, entertainment and social media are less likely to be instrumentally motivated to begin with and also show greater decline over the course of use.

In addition, change over time was different for the types of use. The absolute change in instrumental motivation from the start to the end timing was: 0.0% for productivity, -4.0% for information, -6.0% for communication, -7.9% for entertainment, and -9.8% for social media. For each of the five types of use, we ran a logistic regression analysis with U&G motivation of use as the response variable and timing as the explanatory variable. The decrease from the start to end of use was significant ($p<.001$) for communication, entertainment, and social media, but not for productivity and information. This suggests that when phone use was instrumentally motivated, type of use influenced how likely participants were to switch to a habitual intention.

Our results also suggest two pathways between the U&G motivations and types associated with a loss of sense of agency. In interviews, participants reported mindlessly checking their phone "without even realizing it." This forms a habitual checking pathway in which habitual motivation leads to entertainment and social media use. The erosion of motivation pathway reinforces the same relationship in the other direction. For the entertainment and social media U&G types, participants were most likely to shift towards

habitual motivations later in the course of app use. In other words, as time went on during a session of use, participants reported losing control over goal-oriented, instrumental use.

Micro Escapes

Participants reported that their phone served as a temporary relief from negative states, which we call a micro escape. In this case, participants were not concerned with the meaningfulness of smartphone use, but rather whether or not it helped them cope with undesirable real-life emotions and situations. Participants held different opinions as to whether micro escapes were helpful.

Interview Data

Escaping from Negative States. Participants described turning to their phone during a wide variety of challenging states, both internal (e.g., boredom) and external (e.g., a dispute at work). For instance, a participant who works in customer service shared, “I find my phone to be a way for me to get away and distract myself, instead of sitting there stressed out because I just argued with a customer” (P10). Instead of enduring such undesirable states, smartphones provided participants with an instant way to escape from almost any situation.

Participants turned to their smartphone for relief when they experienced low emotional valence: “I mostly scroll through social media just to get through the day, like I'm bored so I scroll social media. Otherwise, I don't think I scroll social media when I'm happy” (P3). Low valence states that participants said prompted micro escapes included stress, frustration, annoyance, and being upset. One notable exception to this theme was P8, who saw no relationship between his affective state and how he used his phone: “My emotions have absolutely nothing to do with my phone unless I have to make a call that’s of importance.” On the whole, however, participants believed that micro escapes were more common when they were in a negative mood.

The relationship between smartphone escapism and arousal (energy level) was ambiguous. On the one hand, participants reported using their phone to escape when they

were upset, annoyed, and overwhelmed, states that are associated with high arousal ([203]). On the other hand, participants shared that they often used their phone in this way when they were bored, fatigued, and tired, states linked to low arousal [203]. Four participants observed that micro escapes grew more common as they got tired at night.

Participants characterized smartphone use that was prompted by low valence or low arousal as lacking in meaning and purpose:

> *When my mood or when my energy was lower, I tended to reach for my phone to perform tasks that were very nonspecific or did not have any sort of value associated with them. I would either get on Facebook and scroll through my feed or go to YouTube or any activity that had no purpose associated with it other than to pass the time or to distract me from whatever was going on at the moment. (P3)*

Passive consumption was the main pattern of use at these times, as indicated by verbs such as "scrolling," "browsing," "watching," and "checking to pass the time." Social media and entertainment were the types of use that participants most often described as micro escapes. To escape, participants turned to passive consumption that in-of-itself they reported was not very meaningful.

Do People Find Micro Escapes Helpful? Participants held mixed opinions about whether smartphone micro escapes were a helpful response to their negative states. Some participants felt that pulling out their phone helped them cope with certain states (particularly negative emotions with high arousal), whereas others wished they did not do so as often (especially when it replaced a work-related task). Our interview data cover only subjective experience of an individual: what participants themselves viewed as useful (or not). However, we note that objective outcomes (e.g., productivity) and the experience of third parties (i.e., other people affected by an individual's smartphone use) are also important to study.

Micro Escapes for Emotional Self-Regulation. P5 described the phone as a "great distraction tool" that reduced his anxiety: "It takes my mind off of it so that I'm not worried about it, or bothered by it." Four others also found the phone to be an effective tool for emotional self-regulation. P10, who works in customer service, said, "I tried to help them

out and they cursed me out. I'll turn to my phone in order to just pretty much keep me calm and keep me sane pretty much." In cases where participants found micro escapes to be helpful, the emotions they addressed tended to be low in valence, but high in arousal (e.g., upset, stress, frustration, annoyance, and anxiety ([203]). Five of participants reported that smartphone use helped them to cool down these 'hot' emotions.

Other participants felt that smartphone micro escapes were not beneficial, particularly when used to escape low valence and low arousal states (e.g., boredom and fatigue). P6 wished that she did not use her phone when she was bored at night, comparing it against other activities she considered more productive:

> *I definitely noticed that, as the night goes on and as it gets later, I'd look at Facebook more and more. I think it's just because I'm getting tired and bored, and it's just a way to pass the time, when, in reality, I could be actually doing stuff at my house and talking to someone in person and getting things done… I feel like it's a waste of time. (P6)*

Like P6, P3 also turned to her phone for escape when she was bored or mentally taxed:

> *Why do I go on Facebook when I'm bored or I can't write what I want? I don't know… it would have been to just disengage my brain from my task and then I think the ideal is to go back to it with a fresh mind. But I don't think going on Facebook actually accomplishes that. (P3)*

P3 was at a loss to explain why she used her phone to escape when she knew it to be ineffective: "The thing is like I never feel any better, but I still do it. It's such a weird realization. Yeah, the phone usage does not help the situation whatsoever, yet I still do it." This habitual nature of use contrasted against the aforementioned cases where participants described more consciously choosing to use their phone to relieve negative emotions.

Micro Escapes in Social Situations. Four participants valued how smartphones helped them escape from social situations. For P2, who described herself as a "sociable introvert" who sometimes needs "a little escape from the humans," her smartphone provided her with the means to do so: "It's not okay in our society to be like, 'You guys are exhausting. I'm just gonna stand over here for five minutes and take a deep breath, and then I'm gonna

come back to the party or the picnic or whatever.'" P3 similarly liked using his phone to escape from socializing with relatives. Both participants appreciated that they could use their smartphone to escape while in social settings, a form of use they viewed as socially acceptable.

P9 described how his phone helped him to cope with the stress of family life without having to leave the kitchen: "You can't step outside for a minute each time it gets hectic, because it's always hectic with three children. Again, I end up using [apps and games] to uplift myself and relax." For P9, a micro escape was a helpful emotional self-regulation strategy relative to physically removing himself from the setting. We cannot say whether other family members share this opinion.

5.4 — Discussion

Our results reveal systematic ways in which habitual use, type of use, and feelings of meaning all move together. We sought to better understand, first, which uses and gratifications are associated with meaningless experiences, and, second, how these uses and gratifications change as the user engages with an app. Our results show that both the motivation of use (i.e., whether a user picks up the device with intention or out of habit), and the type of use they engage in (as defined by our U&G taxonomy) each independently predict the meaning the user will derive from the experience. Our results also show that habitual use leaves users feeling a lack of control over their own behavior and that these are patterns of engagement that users would like to change.

Characterizing Meaningless Experiences

Reflecting on their phone use generally, participants repeatedly told us they derive relatively little meaning from passively consuming social media and entertainment content. However, they also explained that this was a function of the type of use they engaged in rather than the content itself. Though participants consistently cited Facebook as a source of feelings of meaninglessness, these reports reflected their attitude toward passively scrolling through content, not their attitude toward engaging with their friends

through the platform. This suggests that with respect to the meaning users derive from their phone use, the type of use an individual engages in may be a more useful unit of analysis than app name or app category. Though it can also be valuable to ask questions of particular apps, our results show the value of U&G in this space and that an app-level analysis might miss distinctions specific to the type of use. For example, while some prior work reports that Facebook use predicts declines in subjective well-being [119], other work finds that Facebook use increases social capital and feelings of connectedness [66]. Our results suggest that these seemingly contradictory perspectives may both capture users' experiences accurately and that their differing conclusions may arise from distinctions in the type of use and the gratifications users are seeking.

In addition to the type of use participants engaged in, the motivation behind app use also predicted meaningfulness. Passively browsing social media or consuming entertainment was more meaningful when it was a conscious choice on the part of the user and less meaningful when it was an instinctive reflex. Prior work has shown that "Short-duration, Isolated, Reward-Based" (SIRB) phone use (such as passive social media consumption) is associated with habit-driven phone use, and that these experiences erode the user's intentions [179]). Here, we show that using the phone to engage with SIRB activities and habitually motivated use are each independently associated with less meaningful experiences. Each of these factors independently contributes to a loss of meaning, over and above the contributions of the other.

This is important from a design perspective, because participants reported that they are dissatisfied with their engagement in meaningless activities. In theory, there might be no need for designers to promote experiences that users find meaningful: if users felt good about engaging in habitual phone use and the SIRB activities they characterize as meaningless, one might argue that these are scenarios to support. But consistent with the construct of lagging resistance ([13]) and other prior work on SIRB experiences [94,212], many of the participants in our study reported they would like to change this usage. Prior work on SIRB activities shows that these experiences lead to checking habits and

increased phone use [179]). Here, we further show that these patterns of use predict meaningless experiences that participants wish to reduce.

Lack of Sense of Agency and Meaning

The relationship between habitual use and a lack of meaning persisted across all types of use, even types of use that participants usually found meaningful, like getting things done or connecting with others. Participants shared that they would frequently use their phones without realizing what they were doing and that these experiences lacked meaning. This is consistent with the underlying principles of U&G, which conceives of users as active agents seeking specific gratifications through specific uses [202:69]. Our results suggest that experiences that erode intention and promote habitual use disrupt this core pathway to a user valuing a medium.

These findings imply that even persuasive interfaces designed to draw users into meaningful activities may promote dissatisfaction by pushing users to engage mindlessly. Designers of persuasive interfaces that intend to promote meaningful activities should also consider when and why users might want to engage in this activity of their own volition. A game that encourages a user to solve math problems, care about environmental sustainability, or remember to vote might promote an activity that she finds meaningful, yet could still leave her feeling a loss of agency if the interface promotes habitual phone use, checking habits, and erosion of intentions. A value-sensitive design approach [73] could help designers identify not only what habits users want to form, but also the context-specific ways in which they wish to enact them and the social norms that may help or hinder [137].

Micro Escapes

Some participants explained that they were not always seeking meaning from their phone use; they described sometimes seeking the gratification of escaping from their surroundings and the present moment, not a gratification that comes from phone content directly. This is consistent with prior work in U&G showing escapism as a common

motivation for media use [106,198]. Notably, participants pointed out that these escapes offered them lasting value in times of stress or anger. In these cases, users said that the distraction of the phone allowed them to escape from high-arousal feelings, which dissipated as they engaged in phone activities. It was less clear that micro escapes to avoid boredom or tiredness offered the same lasting value. This provides some suggestion that users may benefit from micro escapes to avoid high-arousal situations but not low-arousal ones. But given that we only encountered this theme through retrospective interviews, the extent to which we can make this claim is limited.

We believe that it would be valuable for future work to examine the lasting impact of micro escapes in both high-arousal and low-arousal contexts. Does turning to Candy Crush in a moment of outrage make it less likely that an individual will fire off an angry email? And does turning to Candy Crush at the end of an exhausting day offer rejuvenation? Participants' reflections suggest that micro escapes can help with the first scenario but offer less assistance in the second. Future work to understand the consequences of escaping in these various scenarios has the potential to offer concrete guidance to users about how they might expect their phone activity to influence their well-being. For developers, it suggests that a general population could benefit from smartphone experiences that detect high-arousal states and assist with emotional self-regulation, as has previously been done with clinical populations [186].

Design Implications

In the technology industry, it's often said that "you build what you measure" [251]. Measuring a user's sense of meaningfulness is undoubtedly harder than tracking their views, taps, or time spent in an app, but can still be done. The overall agreement of the quantitative and qualitative findings in this study suggest that asking for numerical ratings during app use is a valid approach to measuring meaningfulness. However, the interview finding that apps are used as micro escapes from stressful situations shows that qualitative investigation is still critical to understanding the meaningfulness of app use in a broader context. In Facebook's case, the firm setup an offline panel to solicit qualitative input from users and discovered, for instance, that tragic posts were low in engagement

metrics but “really mattered” to people [178] . Developers should consider both quantitative and qualitative data when designing for meaningfulness.

Mobile app designers who want to build meaningful experiences should consider how to respect a user’s sense of agency. To do so, designers should examine how an app supports or distracts a user from their purpose for visiting. Prior work in ubiquitous computing has shown that it is possible to detect or predict when a user is likely to be using the phone mindlessly or problematically [96,221], suggesting that it may be possible for an app to detect when it is and is not being used with intention. If intention is clear, an app could dynamically change its interface to foreground that use case and background others. If a single pattern of use is especially common, a developer may even consider “unbundling" it. For instance, Facebook split off its instant messaging features into a separate Messenger app. If intention is unclear, an app could provide the user with a pause to reflect, rather than defaulting to its most engaging content. For example, when a user opens a new browser window, one could present top news stories or most frequently visited sites, but The Momentum browser extension asks, “What is your main focus for today?”

Apps can also encourage users to move on when their original purpose is achieved. For example, when a user reaches a clean email inbox in Gmail, it shows a blank screen with the message “You’re all done! Please enjoy your day." By contrast, when a user finishes watching a video on YouTube or Netflix, they start auto-playing the next video by default. Designing for a positive disengagement experience could heighten a user’s sense of agency and sense of meaning.

Lastly, smartphone users themselves can also take steps to retain control over their experience. Before picking up their phone, they could form a clear intention for use. Mindfulness-based approaches may aid users to observe and establish psychological distance from the internal triggers that prompt habitual use [126]. Users can also architect the environment on their device to encourage single purpose use, for instance by removing app icons from their home screen and instead accessing them via search

[23,141] . Hiding cues that trigger habitual use could help people keep control and use their smartphone in more meaningful ways [141].

5.5 — Limitations and Future Work

Though we collected ESM and log data from participants, our investigation was exploratory. We used themes from interviews and open-ended ESM responses rather than preconceived hypotheses to guide our analysis. Our claims should be interpreted as qualitative findings rather than hypothesis-confirmation; it would be valuable for future work to evaluate these relationships experimentally. We sampled participants from one cultural context and our sample was over-representative of younger adults, so our results may not generalize to other populations. In future work, analyzing individual app use as part of a larger sequence of use (e.g., use that follows an incoming message/notification) could also yield further insights.

Our findings suggest specific links between motivation of use, type of use, and meaning, and they suggest two pathways—specifically, checking habits and an erosion of purpose—that lead to a loss of sense of agency. A future, large-scale quantitative investigation would help confirm or refute the existence of these links and help to develop these themes into formal theory if applicable. Future work also remains to examine these relationships in the larger social context in which phone use occurs; here we examined only how specific motivations and types of use affect the individual.

U&G is one of many lenses that can be used to understand people's experiences with their smartphones. It has been critiqued for its reliance on self-report [220], its emphasis on the individual without consideration of the collective [239], and its assumption—in contrast to a media effects perspective—that users have full agency when they choose to engage with technology [64]. Despite its shortcomings, U&G has proven over many decades to be one productive way of understanding how individuals' engage with media and technology [202]. Similarly, although we chose to focus on

meaningfulness, concepts such as self-actualization and meaning-making [228] also shed light on eudaimonic well-being. Using other frameworks to examine how people derive meaning from their smartphone use would provide a valuable complement to the data we present here.

Conclusion

In this chapter, we used a U&G perspective to examine how motivation (i.e., a specific intention or a general habit) and type of use (e.g., information-seeking or communication) affect the meaning that a user derives from phone use. We developed a typology of 5 different smartphone uses and gratifications and found that users routinely used the same app for different purposes. Passively browsing social media and consuming entertainment felt less meaningful to participants than other types of use. Separately, some participants said that they do not always seek meaning from their phone; sometimes they seek escape, which can offer lasting value by distracting or calming them during high-arousal emotions (e.g., anger).

Returning to the central theme of sense of agency for time spent on digital interfaces, we found that using one's phone out of habit—for any type of use—reduced meaning. Participants explained that they wanted to cut back on this kind of habitual use, which they associated with a loss of sense of agency. We also found that as a time passed during a session of app use, participants lost control over their original goals for use and slipped into habitual use, leading them to regret how they spent their overall time in the app. Designers who seek to promote experiences that users find meaningful can consider the types of use they promote, the extent to which they erode users' intentions, and how their designs are likely to promote or inhibit sense of agency.

CHAPTER 6 — DESIGNING TO INFLUENCE SENSE OF AGENCY OVER APP USE

In this chapter, I describe our CHI 2021 publication (in order of authorship: myself, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker) in which we address: What design features within apps can we change to support user sense of agency? (**TQ2**) [136]. We first conducted a survey with 120 users of the YouTube mobile app to identify specific features that led them to feel more and less in control over how they spend their time on YouTube. For example, playlists, search, and watch history mostly supported control whereas recommendations, ads, and autoplay tended to undermine control.

We then present our findings from two co-design activities with 13 YouTube users, where we mockup changes to the existing interface that would influence sense of agency. We found that the mechanisms that users preferred changed depending upon the situation. When they had a specific intention in mind, such as watching a tutorial on how to solve a Rubik's cube, they wanted more control. By contrast, when they did not have a specific intention in mind and just wanted to relax or felt bored, they wanted YouTube to take control.

Taken together, our work suggests that digital wellbeing designers might use *adaptable interfaces* that support greater control when the user has a specific intention (e.g., a search-first interface), but provide less control when the user has no specific intention in mind (e.g., a recommendations-first interface). This sets up my work in the next chapter in which I develop, deploy, and evaluate one such adaptable interface, the SwitchTube mobile app.

Introduction

> *At Netflix, we are competing for our customers' time, so our competitors include Snapchat, YouTube, sleep, etc.*" - Reed Hastings, Netflix CEO [240:50]

In the attention economy, social media apps employ a variety of design mechanisms–such as eye-catching notification icons and never-ending autoplay–to maximize their share of the user's time. Designers and content producers often achieve this by intentionally creating experiences that diminish user sense of agency [15]. People self-report that their desire to consume media frequently conflicts with their plans or goals and that they fail to resist about three-quarters of the time [59]. And loss of control is a key component of many measures of problematic technology use [42].

In response, digital wellbeing researchers have innovated what we term *external mechanisms* that help users manage or monitor their app use, such as lockout timers [109] and productivity dashboards [111]. While these mechanisms apply universally to many different apps, they do not change the *internal mechanisms* within an app, such as autoplay, that might lead it to be problematic in the first place.

One promising approach is to redesign these mechanisms for a greater sense of agency, i.e., the feeling of control over one's actions and their outcomes [22]. Low sense of agency over technology use is associated with negative life impacts such as a loss of social opportunities, productivity, and sleep [35] that often motivate digital wellbeing efforts to begin with. Moreover, a lack of sense of agency itself can be understood as a driver of the dissatisfaction that people often feel with their social media use [147].

In this work, we take the mobile app for YouTube, the most widely used social media service in the United States [184], as a test case to understand and redesign how internal mechanisms influence sense of agency. The design of YouTube must balance the interests of many different interested parties. For example, policymakers may wish to exert control over extremist content. Advertisers may wish to control how much time users spend on ads. Designers may wish to control how much time users spend in the app.

Content creators may wish to control how much time users spend on their channel. All of these parties merit consideration, however, in this work we focus specifically on users and how design influences the control they feel over the time they spend in the mobile app.

We investigate two research questions in two studies that build upon each other:

> **RQ1.** *What existing mechanisms in the YouTube mobile app influence sense of agency?* In a survey, we asked 120 YouTube users which mechanisms make them feel most and least in control of how they spend their time in the YouTube mobile app.
>
> **RQ2.** *What changes to these mechanisms might increase sense of agency?* Based on the responses to the survey, we redesigned four internal mechanisms to change user sense of agency in the YouTube app: recommendations, playlists, search, and auto play. In co-design sessions, we then asked 13 YouTube users to sketch changes of their own and evaluate our mockups. We also asked how much control they would prefer to have in different situations.

The two contributions of this work are:

> **(1)** We identify the internal design mechanisms that influence users' sense of agency over how they spend time in the YouTube mobile app and how they might be changed. While some of these mechanisms are expected (e.g., autoplay), others are less so (e.g., playlists) and suggest promising directions for digital wellbeing (e.g., designing to support 'micro plans' that guide behavior within a single session of use).
>
> **(2)** We distinguish when designing for a sense of agency is desirable from when it might actually go against what users want. Participants in our co-design sessions preferred greater control when they had a specific intention for using the app (e.g., to cook a recipe) than when they had a non-specific intention (e.g., to relax), in which case they wanted to let the app take control. We propose ways for designers to navigate this mixed preference for different levels of control at different times.

6.1 — Background and Motivation

Designing to Undermine Sense of Agency

YouTube is an important case for better understanding the design mechanisms of attention capture. YouTube has over two billion monthly users worldwide and is extremely popular in the U.S., where about three-quarters of adults report using YouTube on their smartphone, with 32% using it several times a day, 19% about once per day, and 49% less often [184]. It is also frequently reported as a source of distraction [1], suggesting that it is a good site for the investigation of attention capture dark patterns. In particular, Youtube's algorithmic recommendations merit special consideration as they drive more than 70% of watchtime [225].

Designing to Support Sense of Agency

Reducing screentime in certain apps is a common measure of success in digital wellbeing tools. The two most popular mobile operating systems, Android and iOS, both come pre-installed with tools for the user to track and limit their time in mobile apps. Within the YouTube app itself, there are also features to manage time spent: 'Time watched statistics,' which shows how much time a user has spent on YouTube in each of the last 7 days, and the 'Take a break reminder,' which periodically prompts the user to take a rest. A strength of addressing digital wellbeing via such screentime tools is that time spent is easy to track and easy to understand.

However, a weakness of this approach is that reducing screen time is often a poor proxy for what users actually want. Instead, user intentions are often highly specific, such as wanting to reduce the time spent on targeted features of an app (e.g., on the Facebook newsfeed, but not in Facebook groups) or in certain contexts (e.g., when with family, but not when commuting on the bus) [94,139,145].

Within YouTube, there are two digital wellbeing features that do move beyond time spent controls and offer more granular control. The 'Notifications digest' lets a user bundle push notifications together into a single notification each day, which may reduce the triggers that lead to non-conscious, habitual use [140]. 'Autoplay toggle' lets a user

decide to stop the next video from playing automatically; this may preserve the natural stopping point that comes at the end of the video, a mechanism that has been shown to help users set more deliberate boundaries around use [93]. While the notification digest and the autoplay toggle clearly do more than just track and limit time, it is not immediately clear by what measure of success they might be evaluated.

One promising alternative to the screentime paradigm is to design for *sense of agency* for the time that users spend on digital interfaces, the focus of this dissertation. Sense of agency is a construct that refers to the feeling of control over one's actions and their outcomes [22]. Sense of agency can be broken down into *feelings of agency*, that is, the in-the-moment perception of control, and *judgments of agency,* that is, the post hoc, explicit attribution of an action to the self or other [230]. In the present paper, we focus on the latter, judgments of agency with respect to the time that people spend online.

Prior work has investigated different ways that interfaces can support sense of agency. First, some input modalities seem to support a greater sense of agency than others (e.g., keyboard input versus voice commands) [129]. Second, a system's feedback should match a user's predicted feedback [128]. Third, a study of flight navigation systems found that increasing the level of automation reduced sense of agency [17]. These lessons might be revisited in the domain of digital wellbeing, as how an interface modulates sense of agency may vary with context [128].

Design Mechanisms for Digital Wellbeing

The mechanisms of digital wellbeing interventions can be placed along a spectrum (see **Figure 11**). At one end are external mechanisms that monitor or police apps, such as screentime statistics and lockout timers. A hallmark of an external mechanism is that it functions identically across multiple apps, as in a timer that locks the user out of social media, gaming, and video apps. However, external mechanisms do not significantly change the experience within individual apps.

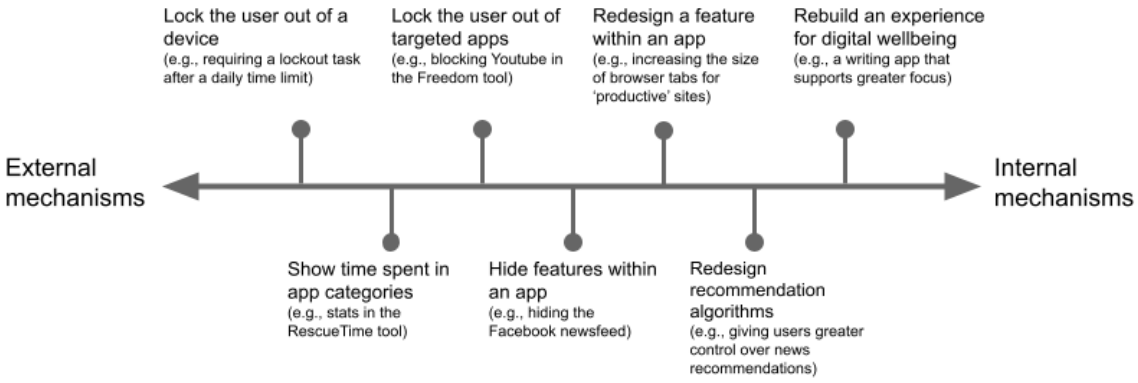


Figure 11. Mechanisms that influence how people spend their time in apps can be placed along a spectrum, as in these examples. External mechanisms monitor or police apps, while internal mechanisms redesign or rebuild the experience within a problematic app. Internal mechanisms offer designers a more targeted way of supporting user agency.

At the other end of the spectrum, internal mechanisms contribute to the redesign or rebuild of an experience. For example, Focus Mode in Microsoft Word redesigns the writing process by hiding all formatting options [6]. Going a step further, the standalone app Flowstate not only offers a minimal interface, but also deletes all text on the page if the user stops writing for longer than seven seconds [227]. Internal mechanisms fundamentally change the experience within a problematic app, or rebuild it into a new experience entirely.

At present, design researchers have innovated many tools on the external side of the spectrum, that monitor and police multiple apps in the same way [47,108,109,163,176]. Likewise, industry designers have built tools that apply the same time lockout mechanism to all apps, such as the screentime tools that come pre-installed on Android and iOS.

In contrast to external mechanisms, the space of internal mechanisms is relatively underexplored (see [86,131][131] for notable exceptions), but holds particular promise for increasing user agency in two ways. First, designers can craft more targeted

interventions with internal mechanisms than with external ones. External mechanisms, such as locking the user out of a device, often require a level of commitment and sacrifice that users might recommend to others, but are reluctant to accept themselves [248]. Whereas an external mechanism might block the Facebook app after time is up, a more internal one could reconfigure the newsfeed to show only content from close personal friends. A redesign of internal mechanisms may be able to remove problematic aspects from an app, while still retaining its benefits.

Second, internal mechanisms shift the focus from fighting distractions to aligning interests. External mechanisms often respond to the temptations of problematic apps with microboundaries [52] or restraints on interactions [182]. However, this sets up an arms race in which the designers of digital wellbeing tools are always in a defensive position. An alternative is for designers to re-envision the internal mechanisms that lead to compulsive use in the first place [233]. Looking at the mechanisms inside of specific apps may encourage designers to not just block existing mechanisms but to innovate better ones, such as Flowstate's seven seconds rule for writing. This paper presents an examination how such internal mechanisms can be redesigned to support sense of agency.

6.2 — Study 1: Survey of 120 YouTube Users

Study 1 examines how existing mechanisms in the YouTube mobile app support or undermine sense of agency (RQ1). We decided to start by investigating user's experiences in the current app before proceeding to design and evaluate potential changes in Study 2 (RQ2). Both studies were approved by the University of Washington's Institutional Review Board.

Participants

Recruitment. To obtain a general sample of users of the YouTube mobile app, we recruited from Amazon Mechanical Turk workers in the United States. Participants were invited to "Help us understand how people spend their time on the YouTube mobile app." They were required to meet four inclusion criteria:

(1) A task approval rating greater than 98% for their prior work on Mechanical Turk, indicating a history of high-quality responses.
(2) Own a smartphone. Three members of our research team tested the YouTube mobile app on both Android and iPhone and found that the app has nearly identical features and only minor stylistic differences, so we accepted users of both types of devices as participants (80 Android, 40 iPhone users).
(3) Spend a minimum of 3 hours on YouTube in the past week (across all devices), according to their time watched statistics in the YouTube app. In the survey, participants saw instructions with screenshots that showed where to find this statistic in the app, confirmed that they had found it, and then entered it into the survey. To see time watched statistics, users must be signed into the app.
(4) Of the time they spend on YouTube, 20% or more is on their smartphone (self-estimated).

Demographics. A total of 120 participants met the inclusion criteria and completed the survey (see demographics in **Table 9**). We excluded responses from an additional 7 participants who started but did not complete the survey. We oversampled men, Asian, and young people relative to the 2019 estimates of the United States Census Bureau [236]. Other participant samples may use the YouTube mobile app differently, e.g., users in emerging countries for whom a smartphone is often their only device for watching videos [224]. Further research is required to determine whether our results apply to other populations.

Table 9. Demographics of the 120 survey participants

Gender identity	Man (63%), Woman (36%), Non-binary (0%), Prefer not to say (1%)
Age range	18-24 (8%), 25-34 (41%), 35-44 (40%), 45-54 (11%), 55+ (1%)
Education	High school (22%), Associate degree (22%), Bachelor's degree (46%), Advanced degree (11%)
Household income (US)	<25K (14%), 25-50K (23%), 50-75K (30%), 75-125K (20%), > 125K (11%), prefer not to say (2%)
Race	(choose one White (69%), Asian (17%), Black (9%), Hispanic/Latino or more) (4%), Native American (2%)

YouTube use. Participants spent a median of 101 minutes per day (interquartile range: 57-156) on YouTube across all devices in the week prior to the survey. Of this time, participants estimated they spent a median of 50% (interquartile range: 30-75%) in the mobile app. For comparison, the YouTube press page states that mobile accounts for over 70% of watchtime [225]. Upon multiplying these two responses together for each participant, we found that participants spent an average of 70 minutes per day in the YouTube mobile app. This is similar to 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 [149]. We neglected to ask whether participants were using the paid YouTube premium service, which removes ads and can play videos offline and in the background; however, Google reports that only 1% of YouTube's monthly visitors subscribe to this service [226].

Procedure

Participants answered questions in an online survey. The initial questions asked about our four inclusion criteria. Eligible participants continued on to background questions about their demographics and YouTube use. The complete survey wording, along with all of the other appendices for this study can be found at: https://osf.io/w3hmd

To investigate RQ1, one question table asked about things that made participants feel most in control of how they spend their time on YouTube (See **Table 10**). A second

question table asked about things that made them feel less in control. The order of these two question tables was randomized. In terms of wording, we chose to ask about feeling "in control," as this is how sense of agency has been measured in previous studies of sense of agency in HCI (e.g., [158]) and on a self-report scale [231]. We used the informal term “things” because, in piloting the survey, we found that testers were unsure about whether certain things (e.g., recommendations and ads) counted as “mechanisms” of the app and we did not want to provide examples that would bias responses. In total, each participant was required to submit 6 responses for things that influenced their sense of agency on YouTube (3 for most in control, 3 for least in control).

Participants were compensated $6.00 for answering all questions, an amount that exceeds the U.S. minimum wage ($7.25 per hour). The survey took a median of 21 minutes to complete (interquartile range: 15-29).

Table 10. The wording and format of the *“more in control”* question in the survey. The example responses here come from a single study participant. All participants also completed a second version of this question table, with the text modified from *“most”* to *“least”* in the Thing Question and from *“more”* to *“less”* in the Explain Question.

	Thing Question: What are 3 things about _____ the mobile app that lead you to feel most in control over how you spend your time on YouTube?	**Explain Question:** How does this thing _____ make you feel more in control of how you spend your time on YouTube?
Thing 1	*“I am able to quickly access my subscribed channels.”*	*“I don’t spend uncontrolled amounts of time browsing through videos that may or may not be related to what I want to watch.”*
Thing 2	*“I am able to get notifications of certain channels or videos getting posted.”*	*“I will know exactly when a new video goes up that I may be interested in watching. This way I am not randomly checking for uploads and spending extra time searching and browsing.”*
Thing 3	*“Screen/watch time.”*	*“I can follow trends and tell when I am spending more time than usual on the app.”*

Coding reliability thematic analysis

We conducted a coding reliability thematic analysis [24,27], in which we first established reliable codes for design mechanisms and then used them to generate themes

that captured shared meanings. We started by iteratively coding the 720 responses (6 per participant). Each "thing" was analyzed as a single response, combining answers to the Thing Question and the Explain Question (i.e., one row in Table 2). In our first pass, two researchers individually reviewed all responses and met to develop initial codes. At this stage, we eliminated 112 responses without any substantive content, e.g., "I can't think of anything else." Of the 112 responses without substance, 55 came from "less in control" and 57 from "more."

We further limited coding to responses that specified a mechanism within the interface of the YouTube mobile app, i.e., something the app's designers could directly change. This included responses such as, "Recommended videos - Being shown recommended videos is like a moth to a light for me," which was coded as 'recommendations'. It excluded responses about situational factors that are largely outside of the control of the designer such as, "I make my own decisions - I am a conscious person who can make decisions on what I do." This eliminated 141 more responses (59 from "less in control" and 82 from "more in control"). Interestingly, "more in control" included 28 responses that we coded as willpower, e.g., "I make my own decisions," with only 1 such response for "less". This suggests a potential self-serving bias wherein in-control behavior is attributed to one's own willpower whereas out-of-control behavior is attributed to external factors [71]. The other responses that we removed were about characteristics of mobile phones (e.g., "The app is easy to access and tempt me on my phone...") and usability issues (e.g., "it crashes on me every other day or so" and "it consumes a lot of battery life") that are not specific to the interface of the YouTube mobile app. After excluding these responses, we continued with coding the 467 responses that referenced a specific design mechanism.

In our second pass, we applied the initial codes to 120 randomly selected responses and met to discuss. Since one mechanism (recommendations) came up more often than all others, we developed three subcodes for how recommendations affected participant experiences on YouTube. After merging similar codes, our codebook consisted of 21 design mechanisms, such as autoplay, playlists, and multiple device sync. In our third pass,

we each independently coded the same 50 randomly selected responses. Interrater reliability was assessed using Cohen's kappa, with κ = 0.73 indicating substantial agreement [122]. In our fourth pass, we each coded half of the remaining responses, discussed the final counts, and selected several representative quotes for each code. The first author then wrote up a draft of the coding results and reviewed together with the other authors. We mapped codes (design mechanisms) to potential themes, generating three higher-level themes that structured our final writeup. In our analysis and writeup, we noted cases where responses for an individual code were split with regards to a theme, e.g., 'notifications' sometimes supported and sometimes undermined 'planning ahead'.

Results and Analysis

Design Mechanisms. 467 responses referenced a specific design mechanism (246 for less in control, 221 for more in control). Nine mechanisms were described as influencing sense of agency 15 or more times and are the focus of our analysis. **Figure 12** provides a glanceable view of how many times each of these nine mechanisms was mentioned as leading to more or less control. **Table 11** shows the same data with a description and example response for each mechanism. Appendix I contains annotated screenshots that show the exact implementation of these nine mechanisms in the YouTube mobile app as they appeared when participants provided their feedback.

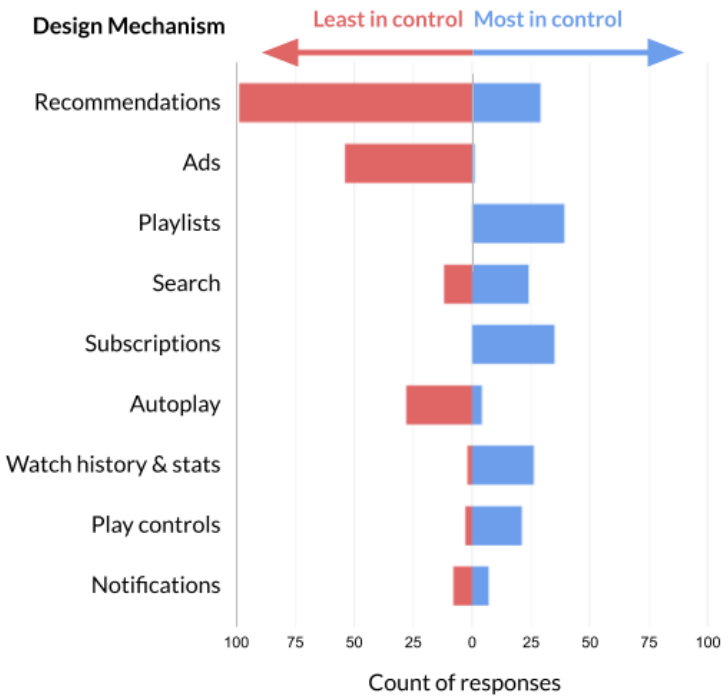


Figure 12. This diverging bar chart shows how many times these nine design mechanisms led participants to feel more control or less control. Features are sorted from top-to-bottom according to which ones were mentioned the most frequently by respondents. Recommendations, ads, and autoplay primarily made respondents feel less in control. Playlists, search, subscriptions, play controls, and watch history & stats primarily made respondents feel more in control. Notifications were sometimes mentioned as leading to more control and sometimes to less.

In summary, recommendations were the most frequently mentioned mechanism, accounting for 27% of all responses. Recommendations, ads, and autoplay primarily made respondents feel less in control. Playlists, search, subscriptions, play controls, and watch history & stats primarily made respondents feel more in control. Notifications were divided with about half of responses in each direction.

Table 11. This table shows nine design mechanisms that were mentioned 15 or more times in response to the survey question: "*What are 3 things about the mobile app that lead you to feel [most | least] in control over how you spend your time on YouTube?*" Design mechanisms are shown in the order of frequency of mention. The most frequently mentioned mechanism, recommendations, is shown with 3 subcodes. The representative quote(s) column shows one typical response for each design mechanism; both a "*more in control*" and a "*less in control*" quote are shown if the minority opinion on the direction of control was more than 20% of total responses.

Design Mechanism	**Description**	**Count of responses**	**Less in control (% of responses)**	**Representative quote(s)** (2 quotes if minority opinion on direction of control >20% of responses)
Recommendations... (see 3 subcodes below)	Recommended videos on the home, explore, & video player screens.	128	77%	See subcodes in the 3 rows below.
/ Irrelevant recommendations	Repetitive, dull, or generic recommendations that the user is not interested in.	42 (of 128)	100%	*"The related videos are sometimes videos I've seen before, over and over."*
/ Relevant recommendations	Engaging or catchy recommendations that the user is interested in.	45 (of 128)	53%	*"YouTube has very good algorithms that know what I like, when I want it." —VS.— "I have a hard time not looking at the suggested videos that the algorithm picks for me... I almost always justify watching just one more video."*
/ Customization settings	Settings to customize location, quantity, or content of recommendations.	41 (of 128)	81%	*"Not having control over the trending list. I feel like I'm force-fed content."*
Ads	Ads that appear before, during, and after videos in the player.	55	98%	*"I feel as if I am forced to watch ads, which can suck up a lot of time."*
Playlists (includes Watch Later)	Creating, saving, and playing a list of videos. Watch Later is a default playlist for all users. Playlists autoplay all videos on the list.	39	0%	*"I can create playlists or queue videos in advance to limit what I watch to a specific list instead of endlessly searching around for what I want."*
Search	Searching for videos.	36	33%	*"Very efficient and relevant searches." —VS.— "Countless videos have nothing to do with my latest search request."*
Subscriptions	Follow specific video creators.	35	0%	*"I can choose the content creators I want to follow so that I can limit my time to specific creators I enjoy the most."*
Autoplay	Automatically plays a new video after the current one. Can be toggled on/off.	32	87%	*"I feel like I have little control whenever YouTube takes it upon itself to just play whatever it feels like playing."*

Watch history & stats	A chronological record of videos watched and time watched stats in YouTube.	28	7%	*"I am able to view EVERYTHING I do in the app. I can keep an eye if I need to change behavior, what type of videos I watch, everything."*
Play controls	Controls to play/pause, seek forward/back, etc.	24	12%	*"I can start, pause and stop content streaming easily, at any time."*
Notifications	System and in-app alerts with new subscription content, recommendations, etc.	15	53%	*"If I especially like a channel I can know about everything they upload as soon as they do." —VS.— "Notifications draw me to YouTube and create my schedule for 20-30 minutes. This creates an addiction."*

How Existing Mechanisms Influence Sense of Agency

The design mechanisms we identified in the YouTube mobile app informed three higher-level themes. First, users experience actions in the app along a spectrum of consent. Second, mechanisms for planning ahead help them feel more in control. Third, the accuracy of YouTube algorithms has mixed consequences for control. The writeup for each theme draws upon examples from our coding of the design mechanisms.

The spectrum of consent. Participants' sense of agency depended on whether it felt like they had 'agreed' to the actions of the app. Participants gave their active consent through actions such as tapping on a play control: "I'm watching a video that's taken too long of my time, so I can just pause it and come back to it. I feel control there." Participants could also issue ongoing consent for the app, e.g., by subscribing to a creator: "My subscriptions show me what I asked to see and I can choose what and when I wish to watch each video." At the other end of the spectrum were mechanisms like autoplay that acted without consent: "It feels weird for the app to start acting before I've told it to do anything." Participants appreciated features such as these that supported their agency to decide what to do next with their time, including when to start and stop an activity.

Non-consent was often felt as a result of (perceived) deception. For example, users disliked ads, but also expected them and indicated their reluctant consent. However, they seemed more upset when the app was unpredictable or violated expectations, as in: "I understand the reason for the ads, but I don't get why some are 5 seconds and you can

skip them while others are 60 seconds and you can't." Other cases where participants felt manipulated included when a "small accidental click" triggered an ad, when video creators were "not upfront" about the products they promoted, and when autoplay "automatically" turned on. Participants disliked when the app openly acted against their interests, but expressed stronger sentiments when they felt that the app also misled them about it.

Planning ahead. Participants felt more in control when they planned their consumption in advance. Playlists helped participants plan how much to watch (e.g., "I can create playlists or queue videos in advance to limit what I watch to a specific list instead of endlessly searching around for what I want"). Participants described the end of a playlist as a "good place to stop", in contrast to browsing recommendations, which they described as "endless." Watch Later, a default playlist on YouTube, also let participants control when and where to watch. A guitar teacher described how Watch Later empowered them to save videos on-the-go and watch them later in their music studio. Watch history & stats also supported planning by providing an awareness that participants could use to adjust their behavior: "I can look at my watch history and see how many videos I have watched today. That puts it into perspective if I should spend time doing something else if I am spending too much time on YouTube." Several participants described using this awareness in conjunction with the Watch Later playlist: "I am able to put a video in my Watch Later playlist if I think I have spent too much time on YouTube for the day."

By contrast, sense of agency was diminished by mechanisms that prompted and pressured participants with suggestions that were hard to decline. Autoplay and recommendations frequently led to this, as in "I often spend more time than I meant to because there is a good related video that seems worth watching so ya know, 'Just one more' which becomes a couple hours." The Watch Later playlist again served as a safety valve in 'just one more' situations: "Watch Later means I don't feel pressured into watching a recommended video from autoplay right when I see it."

Notifications sometimes supported planning and sometimes not. For example, they put participants on the spot: "Based on my viewing history, the app will push me new

content and I may not have the fortitude to not click to view." However, notifications also helped participants plan when to check the app by reducing their fear of missing out: "With notifications I will know exactly when a new video goes up that I may be interested in watching. This way I am not randomly checking for uploads and spending extra time searching and browsing." This may explain why notifications were split between "more in control" and "less in control" responses (47% vs. 53%).

The accuracy of algorithms has mixed consequences for control. Irrelevant recommendations, i.e., those that were repetitive or unrelated to personal interests, universally undermined sense of agency: "Seeing 'recommended' videos that have nothing to do with my viewing history leads to unwanted scrolling and possibly unwanted content." Similarly, irrelevant search results undermined control because they forced participants to keep scrolling for what they wanted, e.g., "I use specific search terms, but I still have to scan past a lot of vaguely or even unrelated stuff to find what I want."

For relevant recommendations, participants' control responses were divided nearly 50-50. In contrast to irrelevant recommendations, relevant ones supported control with their personalization (e.g., "It has some very good algorithms that know what I like, when I want it") or with suggestions that reached just beyond the users' comfort zone (e.g., "I can expand my tastes based on my own preference"). However, relevant recommendations sometimes undermined control by being too engaging, i.e., recommending videos that users watch, but that are unplanned and later regretted. This was captured in participants' use of terms like the "wormhole" (two mentions) and "rabbit hole" (five mentions), as in "The way that videos get promoted to my home page and have appealing thumbnails– I end up clicking on them and wonder how I got to this place and why I am watching this video. I ended up going down the rabbit hole and watching the video and then others like it and so on." Some of these recommendations were described as "clickbait" (six mentions) that misled with content that did not meet expectations and sometimes also violated participants' consent (e.g., by showing "inappropriate content"). More often though, participants seemed to like the content, but felt that it was too much (e.g., "At times there is no escape when I become interested in documentary after

documentary") or not the right time (e.g., "Some of the church videos are addicting and I keep watching them at night"). In other words, even when recommendations were personally relevant in terms of the user's interests, they were often not relevant, or even distracted from, their current task or situation.

Given their mixed experiences with recommendations, participants expressed frustration with the customization settings at their disposal (or lack thereof). Participants lacked the ability to customize the location, quantity, and content of recommendations. Having recommendations on almost every screen led to a loss of control: "It seems like there are video recommendations everywhere. They are obviously in my home feed; they are in the explore menu; and they are under and beside and within other videos. It often takes me down the rabbit hole." Up next recommendations that appear below the current video (and autoplay after it finishes) were specifically mentioned seven times. The "endless" quantity of recommendations also made it hard to stop watching. Finally, participants also wanted to control what content is recommended, particularly when recommended content did not match their aspirations: "There are cases in a particular day where I just want to watch cat videos. But I do not want my entire screen to recommend cat videos." Participants wanted to customize the content of recommendations more directly than just by generating a watch history: "The only thing you can do to control the algorithm is to watch videos. But you get no say how it'll recommend new ones." The vast majority of participants seemed either unaware of YouTube's existing customization settings for recommendations or found them inadequate.

A minority of responses described recommendation settings that do support sense of agency. For instance, three participants appreciated how the settings menu allows them to mark "Not interested" on specific videos, e.g., "When I'm tempted but know a video is not educational I can hide it." In this case, the user is in fact interested in the sense that the video recommendation aligns with their personal interests and triggers their curiosity. However, they must paradoxically mark it as "Not interested" in order to tell the interface to stop showing videos of this kind because they conflict with their longer term goals. YouTube's settings also allow participants to delete videos from their watch history–which

stops them from being used in personalized recommendations–but only one participant mentioned this feature.

6.3 — Study 2: Co-design with YouTube Users

Study 1 identified existing mechanisms in the YouTube mobile app that influence user sense of agency (RQ1). In Study 2, we sought to understand how changes to these design mechanisms might influence sense of agency (RQ2). We conducted 13 study sessions with individual YouTube users that included two co-design activities: 1) sketching participant-generated changes; and 2) evaluating researcher-generated changes that were based on the results of Study 1. Consistent with a research-through-design approach [250], the aim of these activities was not to converge upon a single solution but rather to generate knowledge, i.e., what to design for a sense of agency.

Preparatory Design Work

In preparation for the evaluation co-design activity, five of us on the research team (myself, HZ, JVL, JC, KF), all advanced-degree students in a technology design program, created mockups of changes to mechanisms in the YouTube mobile app that we expected to impact sense of agency. To generate a wide range of possible changes, we started with a design brainstorm that generated 67 different ideas, e.g., creating a 'How-to mode' for viewing only educational content, reducing video playback speed to 50% after a daily time limit is exceeded, or making Watch Later the default action for recommendations. Ideas were reviewed as a group and favorites could be 'claimed' by one author who further refined it. This generated a total of 33 different sketches. We presented, discussed, and then scored these sketches according to three criteria: expected impact on sense of agency (based on the results of Study 1), novelty relative to existing digital wellbeing tools, and feasibility of implementation. Expected effect on sense of agency was weighted twice in our scoring.

We created mockups for the seven sketches with the highest average scores. We wanted participants to evaluate a variety of potential changes to each mechanism, so we

created three versions of each mockup: low, medium, and high-control. For example, the recommendations mechanism in the YouTube app was redesigned to change the number of recommendations shown on the homepage, with the low-control version showing unlimited recommendations, the medium-control version showing only three recommendations with a button to “show more,” and the high-control version not showing any recommendations (see images in **Figure 13**). To focus on RQ2, our results and analysis here address only the four mockups (see **Table 12**) that directly change one of the existing internal mechanisms in YouTube that we identified in Study 1. The other three mockups we tested—activity-goal setting, time-goal setting, and a timer—are more external mechanisms that might apply equally well to other apps. However, we decided to focus this paper on the unique potential of internal mechanisms.

Mockups of the recommendations redesign

Low-control version:
Unlimited recommendations

Medium-control version:
Shows 3 recommendations, then a click-to-show-more button

High-control version:
No recommendations

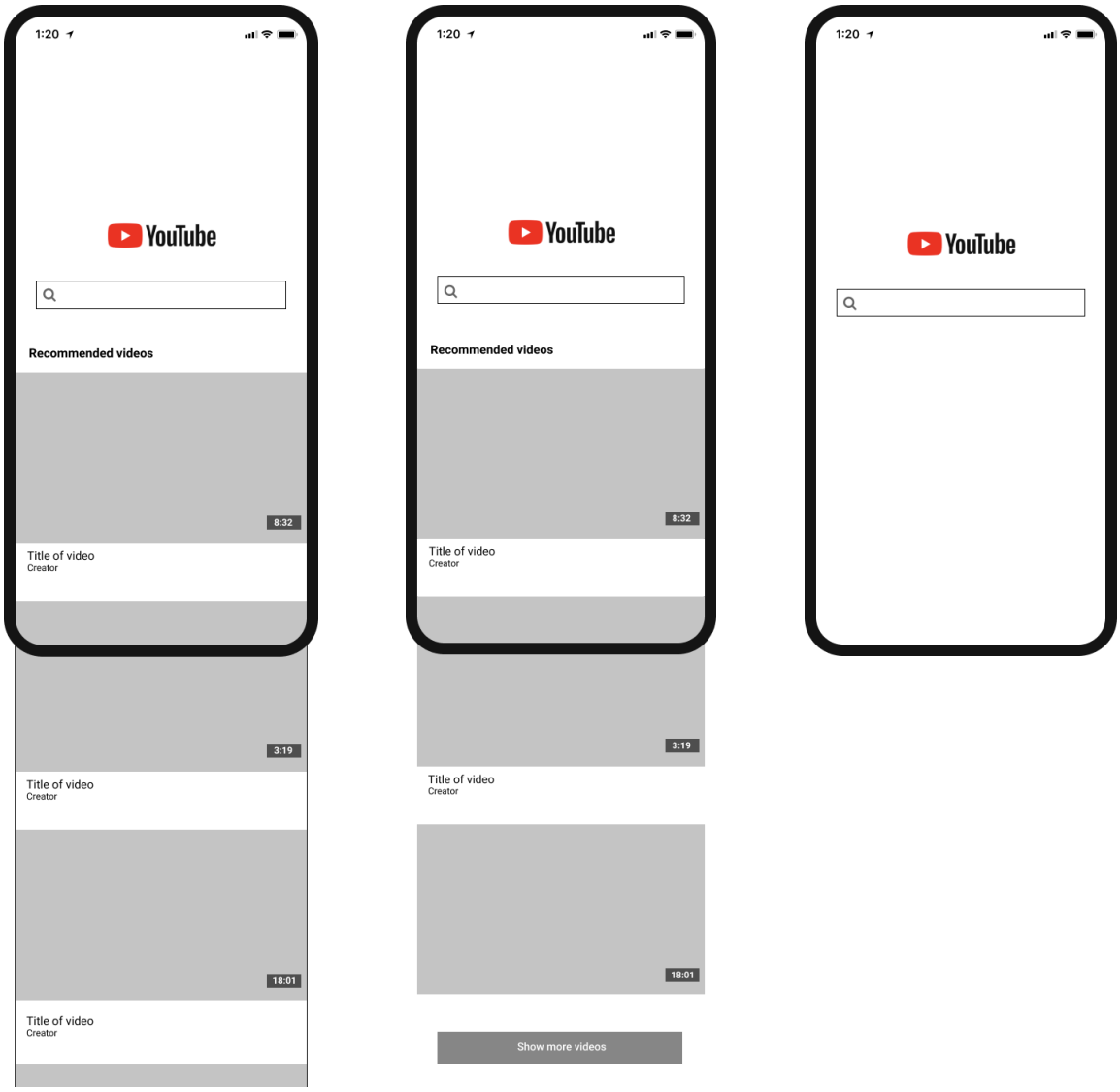


Figure 13. Mockups of the redesign of the recommendations mechanism. We created three versions of the mockup that we expected to offer different levels of control. These 3 versions of each redesign were evaluated by participants in the co-design evaluation activity.

Table 12. This table describes our redesigns of 4 existing mechanisms in the YouTube app. We created three versions of each mockup that we expected to provide different levels of control to the user: low, medium, and high. Appendix II describes more details about the search redesign and the three additional mockups we created, which we do not report on here.

Redesigned mechanism	Dimension of change	Low-control version	Medium-control version	High-control version	Related experience for users (as described by Study 1 participants)	Comparison to current version of YouTube mobile app
Recommen-dations (recs)	Number of video recs on home screen	Unlimited recs	Shows 3 recs, then a click-to-show-more button	No recs	Endless recs often undermine sense of agency	Similar to low-control version
Playlists	Prominence of button to save a video to the Watch later playlist	No Watch Later button	Small Watch later button	Large Watch Later button	Watch Later playlist lets users plan ahead, reduces pressure to watch now	Similar to medium-control version
Search	Prioritization of viral vs. relevant results	Prioritize "viral" results	User can toggle between "viral" & "relevant" results	Prioritize "relevant" results	Sometimes recs and search results that are too engaging undermine sense of agency	Similar to medium-control version
Autoplay	Degree of consent required to play the next rec	Autoplay the next rec	Show the next rec	No next rec	Autoplaying videos without consent undermines sense of agency	Similar to low-control version

We note that although our research focuses at the level of 'design mechanisms,' the details of these designs matter. For instance, although the recommendations in the current version of YouTube seemed to reduce sense of agency in most of the Study 1 responses, a different implementation of 'recommendations' might produce different effects. This is true of our mockups too: in our search redesign we showed a task-oriented example query ("How to cook a turkey"), whereas a leisure-oriented example query (e.g., "Funny cat videos") could have led to different results. We include descriptions of the most relevant details of each of these design mechanisms in the body of the paper, screenshots of their current implementation in the YouTube mobile app in Appendix I, and images of all of our mockups in Appendix II.

Participants

Recruitment. We recruited YouTube users in Seattle via email lists and social media channels to "Help us understand how people spend their time in the YouTube mobile app."

We did not initially set inclusion criteria for participation (beyond adult YouTube users) as we viewed our co-design activities as exploratory. However, after our initial sessions proved insightful for our team of design researchers, we sent a follow-up survey to participants that asked about demographics and YouTube use. Participants were compensated with a $30 voucher.

Demographics and YouTube use. 13 YouTube users (7 women, 6 men) participated in our sessions. The median age was 29 (range: 18-36). Participants reported using YouTube a median of 52 minutes per day (range: 27-70), again based on checking their time watched statistics in the YouTube mobile app. For reference, this amount of time is slightly lower than the average of signed-in YouTube users (60 minutes) [149] and considerably lower than the median of participants in Study 1 (101 minutes).

Procedures

Sessions included an initial think-aloud demonstration of their current YouTube use, followed by sketching and evaluation codesign activities. The median length of a session was 73 minutes (range: 57-105 minutes).

Think-aloud Demonstrations with YouTube App. In a modified version of a think-aloud-protocol [102], the participant opened YouTube on their smartphone and talked us through a typical engagement cycle (how they start and stop use) [233]. Next, they showed and talked us through the mechanisms that made them feel most and least in control of how they spend their time on YouTube.

Co-design Activity 1: Sketching. To elicit participant-generated ideas, we asked participants to sketch over paper mockups of three key screens: home, search, and video player (see **Figure 14**). Each screen represented a minimal version of a video app without recommendations, rather than a direct copy of the current YouTube interface. We chose this minimal version to encourage participants to generate new ideas, rather than to evaluate the existing interface (which we did in Study 1). Participants were handed a pen and a copy of one mockup (e.g., the home screen) and were asked, “What would you

change on this page to feel more in control of how you spend your time on YouTube?" They then received a second copy of the same mockup and were asked to sketch changes that would make them feel "less in control." Each participant created a total of six sketches (two versions of three different screens). As they sketched, participants were asked to explain their thinking [213].

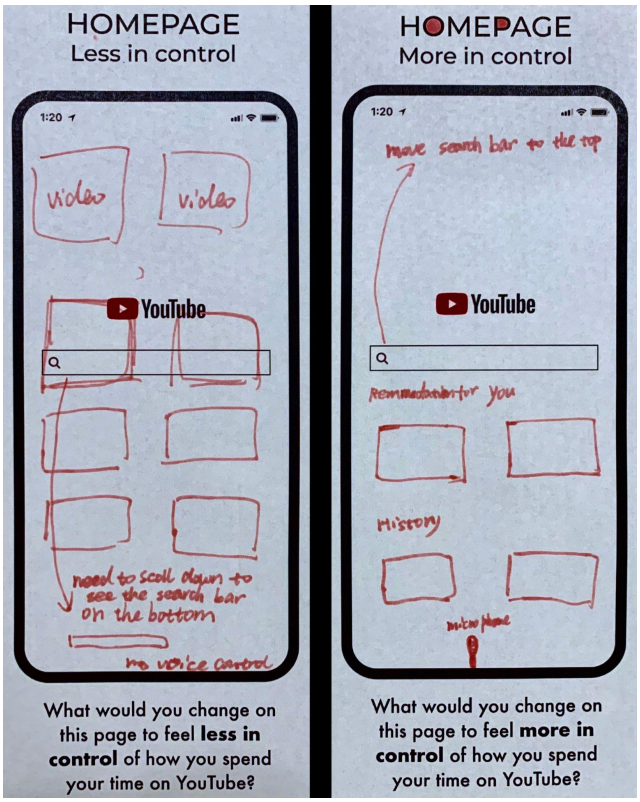


Figure 14. Sketches of the home screen of the YouTube mobile app. The participant (P11) explained that in the "more in control" version, recommendations are based on topics chosen by the user. In the "less in control" version, the user needs to scroll through recommendations to see the search bar at the bottom of the screen.

Co-design Activity 2: Evaluation. To receive feedback on our changes from YouTube users, we asked participants to evaluate our mockups of the redesigned mechanisms in the YouTube mobile app (see Table 5). For each mockup, the three different versions were placed in front of the participant in a random order, they reviewed for about one minute, and then asked any questions they had. We did not tell participants which one was the low, medium, or high control version. The participant was then asked to rank the three

versions in order from the one they would least prefer to use to the one they would most prefer, and explain why.

Codebook Thematic Analysis

We used codebook thematic analysis to analyze the data [25]; [27], wherein we generated themes that are more interpretive than just a summary of all of the data, but less interpretive than in reflexive thematic analysis where the researcher's subject position plays a central role in the analysis [26]. After each co-design session, the researcher leading the session completed a debriefing form with their top three takeaways and shared participant sketches with the rest of the research team. We held weekly meetings to discuss these data and discuss initial ideas. After finishing data collection, all co-design sessions were transcribed. To further familiarize ourselves with the data, three of the authors read the transcripts and again reviewed the sketches. We next independently coded the data using a web app for collaborative coding [223] to generate our set of initial codes. After reviewing this first pass of coding together, we refined and consolidated codes and generated initial themes. Our final set of codes included: user freedom of choice, situational features affecting control, design mechanisms for control, setting clear expectations for the user, and triggers to stop, each of which had further subcodes. We applied our codes to all transcripts and sketches and reviewed the results to create our final themes. For each theme, we extracted vivid exhibits [8], which we used to write analytical memos.

Results and Analysis

We generated two themes about how participants expected changes to the design mechanisms of YouTube would affect their sense of agency. First, participants wanted design mechanisms that provided more control when they had an intention in mind as opposed to when they just wanted to explore. Second, participants envisioned and wanted mechanisms for active and informed choices to increase control.

Specific intentions call for more control. When individual participants reviewed the different versions of their own sketches and our mockups, they were often conflicted about how much control they preferred. It depended upon the situation. When they had a specific intention or goal for their YouTube visit (e.g., to cook a recipe), they wanted design mechanisms that provided greater control. When they had a non-specific intention such as relaxing, they preferred design mechanisms that turned control over to YouTube.

For participants, specific intentions varied from watching a video of a favorite dance, to the latest basketball highlight, to a tutorial on solving a Rubik's Cube. When they had such a specific intention in mind, they wanted greater control than YouTube currently gives them. P4 removed recommendations from their sketch, explaining: "If I have a specific goal, I know what I want, I don't need recommendations to guide my search, I just want to be in control of my search." P2 evaluated our redesign of the search mechanism that emphasized results with higher entertainment value by saying, "I'm probably going to click on it because it's cute and I'm just going to waste so much time. So it's going to make me feel totally out of control of what I actually wanted to come here for." In these cases, participants wanted stronger control mechanisms so that the app would not hijack their specific intention.

Sometimes participants held intentions with a moderate level of specificity, in which case participants wanted to retain some control but also delegate some to YouTube. Often these intentions were topical, as in when P11 wanted to be able to use the app in an "active way" to search and browse videos about programming, but not in a "passive way" to follow just any recommendation. Sometimes, these intentions were temporal, such as when working or studying, participants preferred a version of YouTube that helps them watch a moderate number of videos without making them "fall down a rabbit hole of similar related stuff" (P13). To address these cases, participants sketched both changes to internal mechanisms that were specific to YouTube (e.g., limits on the number of recommended videos) and also more external mechanisms that might apply across a variety of social media apps (e.g., time reminders).

By contrast, when participants had only a non-specific intention (e.g., to unwind or explore), they wanted YouTube to lead the way. Our redesigns of the recommendations mechanism showed either unlimited, limited, or no video recommendations, to which P2 responded "If I came here for a specific reason, like my goal is to learn how-to do something, then I prefer this one without recommendations. However, if I just want to watch something that gets my mind off things, I prefer the one where I can choose to show more recommendations." At times when participants just wanted to be entertained, designing for greater control could actually get in the way. P13 shared, "If you're not giving me recommendations, and if you're making me search, then I'm not in control. Or, I'm in control, but the problem is I'm spending *more time*. There's no point." In other words, even at times when participants desired *less* sense of agency in terms of control over the interface, they still saw the reward in terms of being able to *spend their time online* the way they wanted.

Active and informed choices. The Study 1 theme "Spectrum of consent" addressed whether the user had 'agreed' to an action taken by the app (e.g., autoplaying the next video). To support control, Study 2 participants envisioned more active choices, where the user felt like they were the one to initiate the action. As a step in this direction, P1 described a home screen that presented, "Six categories we think you're most interested in, and then you're at least making the active choice, 'I want to watch some interviews right now.'" In this design, the app's algorithm would recommend a set of personalized topics, but the user would be the one to choose between them. A still more active choice was when the user was the one to generate the set of choices in the first place, as in P7's sketch: "There aren't a billion recommendations on the home screen. It's just a search bar. You go straight to what you want to watch, you watch it, and then you're done." Participants described search as a paragon of user-led choice, and many foregrounded the search option in their sketches to increase control and hid it in ones to decrease control (see Figure 3).

Many sketches also supported more informed choices. These designs made it easier for users to know what to expect from a video by surfacing metadata like view

count, user ratings, and descriptions. Five participants proposed novel metadata, such as an 'activity time' filter that would sort how-to videos by the time it takes to perform the activity they teach, e.g., cook a recipe (P12). Another suggested expert ratings as an indicator of quality (P5). Conversely, in sketches to undermine control, it was common to remove video metadata. P12 likened this to the experience at Costco, a supermarket chain that deliberately shows no signs in its stores [173]: "If you want to go find cookies, they won't actually show you where the cookies are so you literally have to go through every single aisle. You have to go find it."

More choice alone did not lead to more control. In sketches of designs to undermine control, participants covered every corner of the home screen with video recommendations that scrolled infinitely (P11) and in every direction (P5). P13 described, "If they didn't have [recommended videos], it would be a lot harder to follow these different rabbit holes. I imagine that I would have to intentionally seek out another video, so I wouldn't feel sucked in as much." Recommendations prompted a passive form of choice, in which users reacted to the app's infinite scroll of suggestions, rather than making active choices on their own terms.

6.4 — Overall Discussion

Together, our two studies identify design mechanisms that influence sense of agency in the YouTube mobile app and how they might be changed to increase it. In Study 1, participants reported that, in the current app, recommendations, ads, and autoplay mostly led them to feel less in control, whereas playlists, search, subscriptions, play controls, and watch history & stats mostly made them feel more in control. Across all existing mechanisms, participants felt less in control when the app took actions of its own without their consent (e.g., autoplaying a new video recommendation). Recommendations were of special concern and participants expressed frustration at their inability to customize their location, quantity, and content. In contrast, by helping participants plan

ahead for even just a short while, existing mechanisms like playlists and watch stats made participants feel more in control.

When participants envisioned and evaluated changes in Study 2, they wanted more opportunities to make active choices, rather than respond to a set of choices proposed by the app. This preference was stronger when they had a specific intention in mind (e.g., to watch a certain video or topic), whereas when their intention was more general (e.g., to pass the time) they favored turning control over to YouTube.

We expect that our findings on how design mechanisms influence sense of agency on YouTube are most likely to generalize to other social media and media apps where users (a) report feeling out of control at times (e.g., Facebook [147]); and (b) use the app for both specific and non-specific intentions (e.g., Pinterest [42]). We first discuss our findings mostly with respect to our test case of YouTube, before considering implications for digital wellbeing more broadly.

Rethinking What 'Relevance' Means for Recommendations

Recommendations were mentioned by participants as undermining sense of agency far more times than any other design mechanism in the YouTube mobile app, suggesting that recommender systems [193] should be of central concern to digital wellbeing designers. However, they led to a reduced sense of agency via two very different routes: irrelevance and relevance.

First, recommendations were sometimes irrelevant, showing videos in which participants were simply not interested. However, due to rapid advances in artificial intelligence and recommender systems like YouTube specifically (e.g., [50]), one might expect recommendations in social media apps to become more and more relevant in the coming years.

Second, recommendations were sometimes too 'relevant,' which presents a more vexing problem from a digital wellbeing perspective. For example, participants reported

that they sometimes saw too many interesting recommendations (e.g., for documentaries or for church videos late at night), which made them feel a loss of control.

In this case, YouTube's algorithm is arguably too good at a local optimization problem: Out of millions of videos, which one is the user most likely to watch? But it misses a more global optimization problem: Out of many possible actions, which one does the user most want to take? In these cases, recommendations appealed to a users' impulse or short-term desire to watch more videos, but conflicted with their long-term goals, creating a self-control dilemma for the user [63,143].

Our findings call for rethinking what 'relevance' means for recommendations in the context of digital wellbeing. Prior research on recommender systems has argued that "being accurate is not enough," as a fixation on accuracy can lead designers to ignore important facets of user experience like serendipity [155:1]. For participants in our study, sense of agency was clearly a neglected facet of user experience, as YouTube's recommendations led them to actions (i.e., watching more videos) they did not feel in control of. To be clear, this does not mean that Google or others should try to create an 'algorithm for life' that recommends between watching another video, writing a term paper, and going to sleep.

However, it does suggest that recommender systems could first start with the global problem of when to show recommendations, before moving on to the local problem of which items to recommend. For example, a decision not to show recommendations might be informed by the time of day (e.g., 2am is too late), screentime preferences (e.g., when the user has already exceeded their goal of 30-minutes per day on entertainment apps), or explicit user preferences (e.g., only show three recommendations unless I click-to-show-more). In HCI research, sometimes the implication of a user needs assessment is not to design technology, as a new technology might not be appropriate in the context of the larger situation [14]. Similarly, for recommender systems, our findings suggest that sometimes the implication is not to recommend. Prior work has addressed how a system can display the level of confidence it has in its recommendations to the user

[154], but this should be preceded by a more fundamental question of whether or not to show recommendations in the first place.

Whereas both of the studies in this work elicit user preferences ("what users say"), the dominant paradigm of recommender systems today, including YouTube, is behaviorism: recommendations largely neglect explicit preferences and instead rely on behavior traces ("what users do") [65]. The present bias effect [175] predicts that actual behavior will favor the choice that offers immediate rewards at the expense of long-term goals. In this way, recommender systems reinforce the sometimes problematic behavior of the current self rather than helping people realize their 'aspirational self' that reflects long-term goals [65]; [140].

Participants also wanted to customize the content of recommendations, e.g., "I do not want my entire screen to recommend cat videos." Today, the dominant paradigm of recommender systems, including YouTube, is behaviorism: recommendations rely on behavior traces ("what users do") and largely neglect explicit preferences ("what users say"). In this way, recommender systems reinforce the sometimes problematic behavior of the current self rather than helping people realize their 'aspirational self' that reflects long-term goals [65,140]. Designers might address this by making it easier for users to (a) explicitly state preferences for topics they would like to see or not see; (b) explicitly rate recommendations (e.g., show me more like this one); (c) edit their viewing history to influence future recommendations (e.g., delete all cat videos); or (d) select an algorithmic personae to curate their recommendations (e.g., "The Diplomat," who brings news videos from the other side) [86:72]. The current YouTube app offers limited support for these first three features (e.g., users can select from among topics for recommendations on the home page of the app), but participants in our study seemed mostly either unaware of these customization settings or found them to be inadequate.

To summarize, we encourage designers of recommender systems to think beyond just optimizing for the item that is most likely to be clicked, watched, or liked. This includes considering when to show recommendations in the first place. It also means exploring how

recommendations can support user aspirations rather than just reinforce current behaviors, which requires better measures of long-term preferences. Designers and researchers should continue to explore how to personalize recommendations to satisfy these broader user needs, or provide customization options that put users in control - at least to the extent they want.

Designing to Support Microplanning

Behavior change researchers have long known that plans can help bridge the gap between intentions and behavior. In this work, plans are usually crafted in advance through careful deliberation and guide behavior for some time into the future [3]. For example, a screentime tool in this mold might ask the user to review and reflect upon their past usage data and develop a plan for their use over the next month. Participants in our study also 'planned', but they did so in a more ad hoc manner. For example, they queued videos in advance to limit what they watched during a single session or glanced at their Time watched statistics to know whether to watch another video or add it to their Watch Later playlist.

These types of actions might be called 'microplanning,' making lightweight plans that guide behavior for a short time, usually just a single session of use. Our naming takes inspiration from Cox et al.'s coining of the term 'microboundary' to describe "a small obstacle prior to an interaction that prevents us rushing from one context to another," which serves as a 'micro' version of a commitment device that prevents the user from "acting hastily and regretting it later" [52] . 'Microboundary' has helped center an important concept from behavioral economics, commitment devices that restrict future choices to reflect long-term goals [29,210], in the research and development of digital wellbeing tools, e.g., [108,143,187].

Similarly, we hope that the concept of 'microplans' encourages the use of behavior planning knowledge in the design of digital wellbeing tools. For example, this literature finds that plans are more likely to succeed if they specify where, when, and how a behavior will be enacted [77]. A microplan might incorporate just the where part, and be supported

by a video playlist that is tied to a specific location, e.g., song tutorials for my guitar studio. Triggers are also a key component of effective plans [70], so in this case the playlist might be the primary recommendation in the app anytime the user is within 50 meters of the studio. In another example, Hiniker et al. adapted an evidence-based Plan-Do-Review sequence [68] for an app that asked children to plan out their video-watching, finding that it helped them transition to their next activity with ease [95]. In the domain of impulse buying [165], an e-commerce site (or third-party extension) might foreground 'shopping list' tools to support intentional buying.

Different Levels of Control for Ritualized and Instrumental Use

In Study 2, participants suggested ways that the YouTube mobile app might be redesigned to increase sense of agency (e.g., by reducing the number of recommendations it displays). However, such changes might lead to adverse effects as there were also times when participants preferred low-control features. Although HCI design guidelines advise supporting user sense of agency [171,222], we should not assume that a greater sense of agency is always desirable.

Specifically, participants preferred higher-control mechanisms when they had a specific intention in mind and lower-control ones when they had a non-specific intention. This finding broadly aligns with two types of viewing that have been identified in uses and gratifications research on television use [199]: (1) ritualized use, open-ended use to gratify diversionary needs; and (2) instrumental use, goal-directed use to gratify informational needs. On this view, the current version of the YouTube app appears to offer good support for ritualized use, but poor support for instrumental use, as participants often felt that their specific intentions were hijacked by its autoplay and endless recommendations.

How might a single app support sense of agency for both ritualized and instrumental use? One approach is a customizable interface that lets the user switch between low and high levels of control. This can be done at the app-level, e.g., switching between an Explore Mode and a Focus Mode. Or it can be done at a mechanism-level, e.g., YouTube currently offers an on/off toggle for autoplay, but does not provide any way to

toggle recommendations, which were the mechanism most frequently mentioned as leading to a loss of control in Study 1. This approach may be particularly suitable for power users, as prior research indicates that power users prefer interfaces that are customizable (user-tailored) by a toggle, whereas non-power users prefer ones that are personalized (system-tailored) for them [229].

A second approach then is an interface that is personalized for the user based on a prediction model. Recent work has found that classifiers can be trained to predict these types of media use with high confidence, e.g., for Pinterest [43] and smartphone use [96]. For example, if YouTube expects that the user is visiting for ritualistic use, it could remain as is, or even go further to take control as in its Leanback mode for "effortless viewing" that autoplays a never-ending stream of high-definition recommendations [78]. Both our own findings on autoplay and previous work suggest that such a high level of automation would reduce sense of agency [17], but it may still be the interface that the user prefers in this situation. Conversely, if YouTube has high confidence that the user is visiting for instrumental use, it could present a search-only interface and hide all recommendations. Finally, if it has low confidence in its prediction, it could present a middle-ground interface that shows limited recommendations, or it might err on the side of caution and lead with a search-first interface in case the user has an intention to express.

Towards a Language of Attention Capture Dark Patterns

Our findings address what and when to design to increase sense of agency. However, in the attention economy, what might motivate key parties to support such designs? One step is for the design community to develop a common language of attention capture dark patterns that recognizes designs that lead to attentional harms.

Developing such a lingua franca of attention capture design patterns could be integrated into design education [79], influence designer thinking, and reputations, as is done by the name-and-shame campaign of the darkpatterns.org website [28]. At the company level, it could help inspire products that are aware of the user's sense of agency. For example, in spite of the incentives of the attention economy, Apple is now working to

make privacy a selling point [85], e.g., by preventing developers from tracking users across multiple apps without their active consent [5]. At the regulatory level, a recent review of dark patterns by Naraynan et al. notes that if the design community does not self-regulate by setting standards for itself, it may be regulated by more onerous standards set by others [168]. The U.S. Senate is currently considering how to regulate social media, with one bill that would make it illegal to "manipulate a user interface with the purpose or substantial effect of obscuring, subverting, or impairing user autonomy" [153] and another that would ban autoplay and infinite scroll [41]. For designers, the language of dark patterns is an important way to contribute to a broader critical discussion of design practices in the technology industry [79].

We caution that the message of attention capture dark patterns should not be "never X," but rather "be careful when X." Participants in both of our studies reported mixed experiences with many design mechanisms, including autoplay and recommendations. An outright ban on these mechanisms is likely to reduce sense of agency in a substantial number of situations where the user just wants to explore. Instead, a nuanced guide to dark patterns might present examples of the problem, followed by counterexamples where such a pattern is appropriate. While this creates a murky gray middle, it also better describes the effects of the design mechanisms that we identified in our studies.

Limitations

In addition to the previously stated limitations of our participant sampling and focus on design mechanisms as a unit of analysis, our work also has at least two conceptual limitations that could be explored in future work. First, both of our studies asked participants to share their preferences, however present bias [174] predicts that actual behavior will favor the choice that offers immediate rewards at the expense of long-term goals. An in-situ study of how people respond to redesigns intended to influence sense of agency would yield results on ("what users do"), which might need to be reconciled with the present results on ("what users say"). Second, time and attention are

not the only factors that influence sense of agency. By asking participants in both studies to reflect on "...in control of how you spend your time on YouTube" we discouraged participants from considering other factors such as privacy. In Study 2, this may have primed participants to focus on sense of agency over other factors when evaluating which version of the mockup they preferred.

Conclusion

Whereas a common approach to digital wellbeing is designing to reduce screentime, this work takes an alternative approach of designing to increase sense of agency. In two studies, we identify mechanisms within the YouTube mobile app that participants report influence their sense of agency and how they want to change them. We find that participants generally prefer mechanisms like autoplay and recommendations to be redesigned for a greater sense of agency than the YouTube mobile app currently provides. However, there are still times at which participants prefer to lean back and for YouTube to take control. In Chapter 7, I explore how a system can support different levels of agency for different use cases.

For digital wellbeing designers, we highlight a need for recommender systems that better reflect user aspirations rather than just reinforce their current behavior. We also propose mechanisms that support 'microplanning,' making lightweight plans to guide a single session of use, to increase user sense of agency. Finally, we propose language that the design community might adopt to recognize design patterns that impose attentional harms upon the user.

CHAPTER 7 — SWITCHTUBE: TOGGLING BETWEEN EXPLORE AND FOCUS MODE WHEN WATCHING YOUTUBE VIDEOS

This chapter builds on the survey and co-design work with YouTube users described in the previous chapter (Chapter 6) to develop and evaluate SwitchTube, a mobile client for watching YouTube videos. We (myself, Ulrik Lyngs, Karina Shirokova, Raveena Rao, Larry Tian, Himanshu Zade, Sean A. Munson, and Alexis Hiniker) explore how an adaptable commitment interface (ACI)—an interface that commits the user to avoid distractions but can also easily be toggled off—affects user experience. This work addresses my third thesis question: For apps that lead users to feel a loss of control, can alternative versions support sense of agency in-the-wild? (**TQ3**)

We first surveyed 606 YouTube users from the U.S. to learn what changes, if any, they would like to make to their use of the mobile app and what stage of the behavior change process they are currently at. In our second study, we designed and built three versions of the SwitchTube app, each of which has different features intended to support different levels of sense of agency. We then deployed SwitchTube in a three-week field study with 46 U.S. participants and conducted data-driven retrospective interviews with 16 of them. Our SwitchTube work contributes (1) an empirical understanding of the relationship between sense of agency, short-term satisfaction, long-term goals, and time spent; and (2) lessons for adaptable interfaces that support different levels of sense of agency. We plan to submit this work to CHI 2023.

7.1 — Background and Motivation

YouTube as a Site of Problematic Technology Use

People report that the YouTube mobile app is a source of entertainment and education, but also of temptations that conflict with their preferences and goals. For instance, users express concern that they get sucked down the rabbit hole and watch

more videos than they originally planned (Chapter 6). As another example, the RegretsReporter is a browser extension and crowdsourced research project by the Mozilla Foundation to understand the harms of the YouTube algorithm [255]. Project contributors report that they have received unwanted recommendations for videos such as “Blacks in Power Don’t Empower Blacks,” which promote racist language and ideas and “Woody’s Got Wood,” a hypersexualized parody of the animated children’s film “Toy Story.” YouTube is second only to Facebook as a target for intervention among digital self-control tools [143]. These examples illustrate that users have concerns over both the quantity and the quality of the content that they consume on YouTube.

In a study of television consumption [81], 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 13**). They distinguish between four different media effects:

1. **Time choice effects:** the impact of the quantity of television viewing on short-term satisfaction (e.g., regret for lost time, sense of passivity)
2. **Lifestyle effects:** the impact of the quantity of television viewing on long-term wellbeing (e.g., fewer social relations, sleep problems)
3. **Content choice effects:** the impact of the quality of television content on short-term satisfaction (e.g., dissatisfaction with the content of the show they watched relative to other shows)
4. **Cultivation effects:** the impact of the quality of television content on long-term wellbeing (e.g., fostering increased materialism or political extremism)

By limiting time on phones or in certain apps but not the content within those apps, designers of current screen time tools have largely prioritized the management of time choice effects. In the long-term, such choices about how much to consume have lifestyle effects, as in when participants describe struggling to finish a term paper due to digital distractions [145]. Yet cases such as the racist and sexualized YouTube videos

demonstrate that the quality of media content is also a concern for users (content choice effects). In the long-term, these may have cultivation effects wherein a person may find their own character has changed in a direction they regret, e.g., becoming a conspiracy theorist after following ever more extreme YouTube videos [195]. A digital wellbeing tool for YouTube then should first assess not just how to support user goals for quantity of time, but also to support their goals for the quality of content.

Table 13. A taxonomy of the effects of media consumption. Adapted from [81].

	Satisfaction (short-term)	**Wellbeing** (long-term)
Quantity of TV consumption	Time choice effects	Lifestyle effects
Quality of TV consumption	Content choice effects	Cultivation effects

One challenge is that the features that users prefer in digital self-control tools depend upon personality, gender, and culture [151,152]. In my own work (Chapter 6), I found that even the same YouTube users have different needs in different use cases [136]. For instance, when users have a specific intention in mind, they tend 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. One potential solution then is to adapt the user interface to support the varying levels of control desired by the user in that particular situation.

Adapting the User Interface for Different Levels of Sense of Agency

Adaptable interfaces can be tailored to the individual and their situation. 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 (see **Figure 15**). 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.

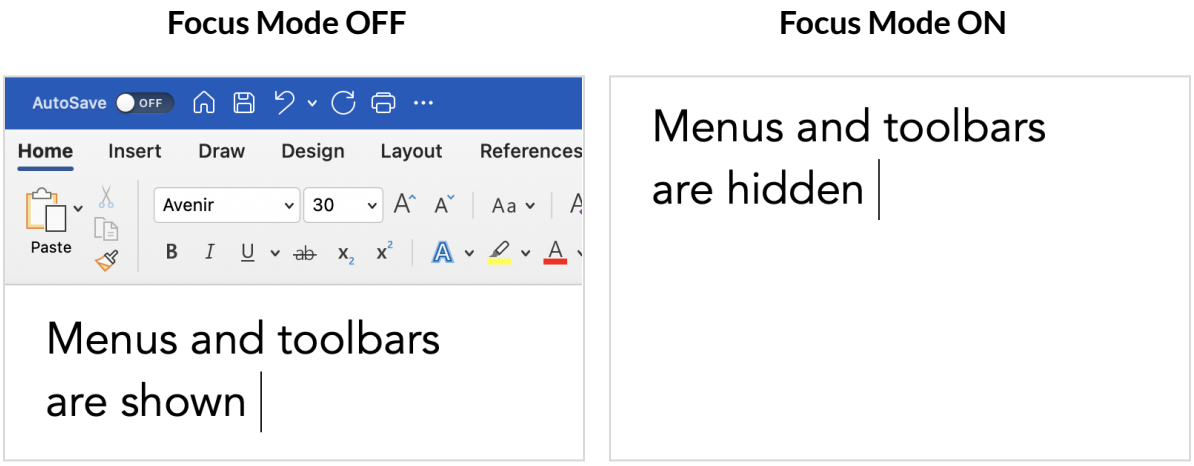


Figure 15. The Microsoft Word desktop app enables the user to manually turn on "Focus Mode," which hides the menus and toolbar.

So how might a single app support different levels of sense of agency for the user? One approach is an adaptable interface that lets the user manually switch between versions designed to support different levels of sense of control. This can be done at the app-level, e.g., switching between an Explore Mode and a Focus Mode. Or it can be at a mechanism-level, e.g., YouTube currently offers an on/off toggle for autoplay. This manual approach may be particularly suitable for power users, as prior research indicates that power users of technology (defined by high usage, expertise, and dependence) prefer interfaces that they can customize themselves (user-tailored). Non-power users, however, prefer ones that are automatically personalized for the user (system-tailored) [229].

Adaptable Interfaces: Undermining the Power of a Commitment Device?

There is a particular unresolved challenge for adaptable interfaces in the domain of digital self-control. Digital self-control tools often function as commitment devices: the user commits to a goal and the tool holds them to it [134]. That is, the present self engages in reflective system thinking and anticipates that their future self, when engaged in automatic system thinking, will be unable to stick with their goal. So the present self delegates their behavioral goals to a digital self-control tool (**Figure 16**). 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. Yet when the user can easily toggle between interfaces, it raises the question how strongly must a commitment device be enforced in order to be effective?

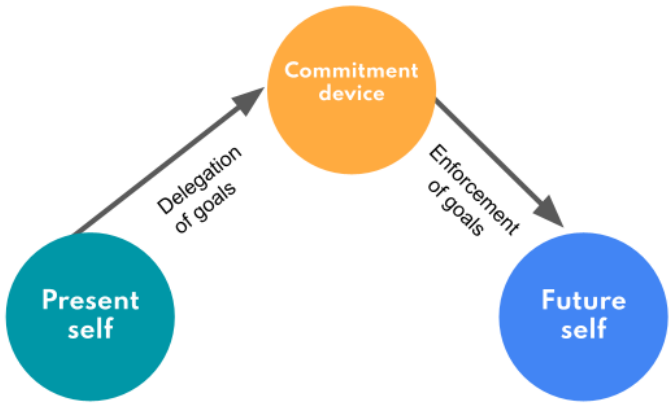
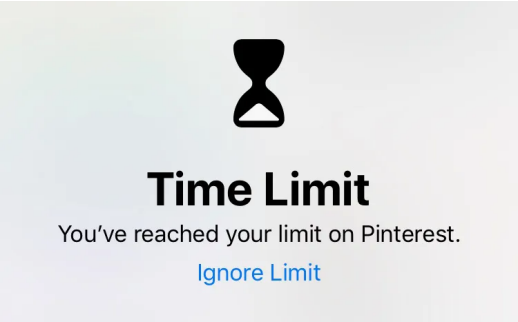


Figure 16. The present self does not trust the future self to stick to their goal, so they delegate their goal to a digital self-control tool which acts as a commitment device. The tool enforces the goals upon the future self. But if the enforcement is too easy to override, will the commitment still be effective? Reprinted from [134].

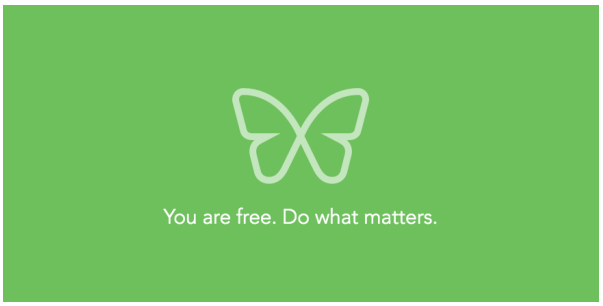
Prior work has found that digital self-control tools are generally more effective at curbing screen time when the severity of enforcement or the degree of friction of the tool is stronger [52,108], an intuition that users themselves hold [234,246]. Too weak and it might be too easy for the user to circumvent their original goal, for example by clicking "Ignore Limit" on a warning that time is up (Figure 2, panel a). 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 2, panel b). This motivates a search for a "Goldilocks" level of enforcement that strikes the right balance: providing the right amount of goal support without leading to psychological reactance against the tool [142].

(a) Weak enforcement
Easily tolerated by users, but goals are also easily and frequently circumvented



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

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



The desktop software Freedom blocks websites without an override option

This suggests competing hypotheses for when an adaptable interface serves as a commitment device, which I shall call an adaptable commitment interface (ACI). 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). 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 will still function as an effective commitment device.

Research Questions

This chapter considers three research questions regarding YouTube use and adaptable commitment interfaces:

- **RQ1: How prevalent is the desire to change the quantity and quality of use among YouTube users in the U.S.?** I draw upon the stages of change model of motivation [188] to understand how and when people want to change their YouTube use.

- **RQ2: When and why do people switch between interfaces in an ACI?** Since users have the choice to switch at any time, will they switch in ways that undermine their commitment? Or will they switch in ways that meet their situational needs?
- **RQ3: How does an ACI influence user experience?** In particular, how does an ACI for digital wellbeing influence measures of sense of agency, satisfaction, and goal alignment among users who want to make a change to their YouTube use?

I 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 ACI designed to answer RQ2 and RQ3.

7.2 — Survey: Stages of Change for YouTube Users

In behavior change, the transtheoretical model (stages of change) is often used to assess an individual's readiness to make a change [188], ranging from pre-contemplation (no intention to change behavior) to action (actively trying to modify behaviors). I draw upon this model to answer RQ1: How prevalent and at what stage is the desire to change the quantity and quality of use among YouTube users in the U.S.? A large-scale survey could assess how commonly users are motivated to change by digital wellbeing concerns, in the specific case of YouTube use, as well as the nature of their concerns. This information could be used to inform the design and development of tools to modify the YouTube experience, as we did with our own SwitchTube app.

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 tell us about their YouTube watching habits and preferences without any mention of changing or reducing use. They 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 14**). We excluded responses from an additional 90 participants who started but did not complete the survey. We oversampled young people relative to the 2019 estimates of the United States Census Bureau [236].

Table 14. 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[4]). Of this time, participants estimated they spent a median of 52% (interquartile range: 20-80%) in the mobile app. For comparison, the YouTube press page states that mobile accounts for over 70% of watchtime [225]. Upon multiplying these two responses together for each participant, we found that participants spent an average of 34 minutes per day in the YouTube mobile app. This is somewhat 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 [149].

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

Procedure

Participants answered questions in an online survey. The initial questions asked about our four inclusion criteria. Eligible participants continued on to background questions about their demographics, technology use, and YouTube use specifically.

To investigate **RQ1**, we adapted a questionnaire from prior behavior change research [243] to assess participants' stage of change with regards to their YouTube use (**Table 15**). 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,"* they answered 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 U.S. minimum wage ($7.25 per hour). This research was approved by the University of Washington IRB.

Coding Reliability Thematic Analysis

We conducted a coding reliability thematic analysis [27] to identify commonalities in participants' 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 [81]. 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 the same 50 responses. The interrater reliability for our codes and subcodes ranged from Cohen's Kappa of 0.66 (substantial

agreement) to 1 (perfect agreement) [122]. The same two authors proceeded to code the remainder of the responses.

Results and Analysis

Stages of change. Based on the survey participants' responses (Table 15), 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. Overall, about half of our sample of general YouTube users were either actively trying or in preparation to make a change to take more control of their YouTube use.

The concept of "lagging resistance" documented for Facebook users [13] 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 there are fewer social pressures to engage in YouTube use, whereas Facebook use is often connected to external constraints such as a fear of missing out on social events. That about half of participants were in the action or preparation stage aligns with previous findings of an underlying discontent with social media use [123].

Table 15. 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 →	Action	40%
No or Not sure >> (go to the next question)		
Do you plan to [...] in the **next month**?		
Yes →	Preparation	8%
No or Not sure >> (go to the next question)		
Do you plan to [...] in the **next six months**?		
Yes →	Contemplation	4%
No or Not sure >> (go to the next question)		
Do you **currently** feel in control of how you spend time on YouTube?		
Yes →	Maintenance	45%
No or Not sure →	Precontemplation	3%

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 participants in the maintenance stage (45% of participants) who already felt in control of the time they spent on YouTube, there was often still a desire to change something about their YouTube use.

Participants shared 817 goals and wanted to change both the quantity and quality of time that they spent on YouTube (**Table 16**). 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 starting a channel, re-organizing saved videos, and reducing data usage.

Table 16. Our coding of the changes to YouTube use that survey participants wanted to make. *Subcodes that accounted for less than 5% of the parent code are not displayed.

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%		

Of note is that the default system screentime tools (Apple Screen Time and Google's Digital Wellbeing) currently address only about 40% of our participant's change goals. 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 my previous research found that users feel frustration over their inability to customize the quantity, content, and placement of recommendations that appear in the interface (Chapter 6).

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 our next phase of our study, the field experiment. In therapy, the stages of change are often used to match the therapeutic process to the stage of the client. 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.

7.3 — SwitchTube Field Experiment: Evaluating an Adaptable Commitment Interface

To answer our two remaining research questions, colleagues and I designed, built, and deployed SwitchTube, a mobile app that manipulates the sense of agency that users have when watching YouTube videos (**Figure 17**). In addition to a low-control (Explore Version) and high-control (Focus Version), **we also created a Switch Version that lets users toggle between low-control and high-control**. This option was based on our findings in prior research that participants wanted interfaces that adapted to their intentions for use (Chapter 6). By comparing these three interfaces against each other, we aimed to understand when and why people switch between interfaces in an ACI (**RQ2**) and the influence of an ACI on user experience (**RQ3**).

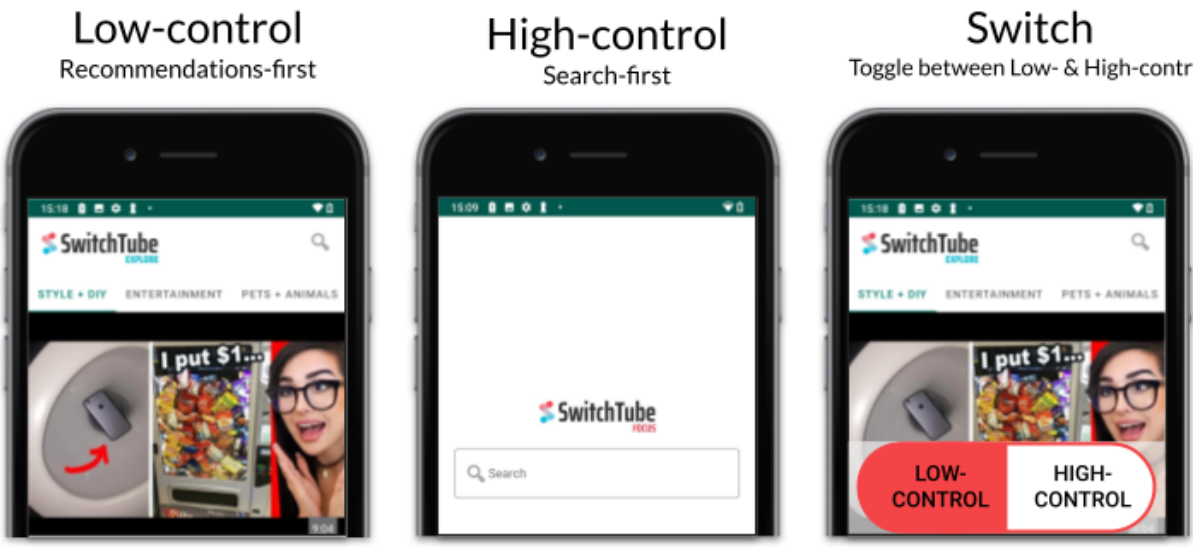


Figure 17. Colleagues and I built and compared three different versions of the SwitchTube app, which manipulate the level of control users have when watching YouTube videos. The Low-control (Explore Version) has a recommendations-first homepage whereas the High-control (Focus Version) has a search-first homepage. The Switch Version is an adaptable commitment interface that lets users toggle between these two interfaces.

We tested the three versions of SwitchTube against each other in a 3-week within-subjects experiment [132] with 46 participants, all YouTube users who were actively or imminently trying to change their use. We collected data from logs of app use, experience sampling, and exit interviews. We found that in the case of SwitchTube, an ACI provides the best of both worlds: participants experienced higher sense of agency, satisfaction, and personal goal alignment when they could toggle between multiple interfaces to suit their situational needs.

Preparatory Design Work

Based on our findings in prior research and our survey in this work, we set out to understand how to design a mobile app that provides users with a greater sense of agency over their video watching experience. Together with a team of 6 other advanced-degree students in a technology design program, I went through a design process of ideating, prototyping, building, and piloting before we tested SwitchTube in the field.

Ideating. We started by ideating 10 different design dimensions to manipulate user sense of agency over time spent based on our own work in Chapter 6 and a few attention capture dark patterns that have been proposed in the literature (e.g., "playing by appointment" in [247]). For example, the Time Pressure dimension ranged on a spectrum from no control to full control and the Content Selection dimension 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 had multiple feature sets). For example, for Time Pressure, we imagined video recommendations that expired if not watched within 30 minutes (low sense of agency), ones that expired within a week (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 impact, novelty, and technical feasibility.

Prototyping. We created paper mockups for the 3 highest-scoring feature sets and evaluated them in 13 co-design sessions with YouTube users as described in the previous chapter. For example, **Figure 18** shows our prototype for Content Selection with three different versions of search results. Our initial plan had been to build three different versions of SwitchTube, each of which supported a different level of sense of agency (low, medium, high) to find a "Goldilocks" level of control as has been explored in prior work on lockout mechanisms [108]. However, one of the relevant findings in our co-design was that rather than having an absolute preference, participants wanted different levels of control for different situations. When they had a specific intention in mind, they preferred a search-first interface, whereas when they just wanted to relax or pass the time, they preferred a recommendations-first interface. This led us to revise our initial plan and

instead design a low agency version, a high agency version, and one version where users could switch between the two (hence the name SwitchTube).

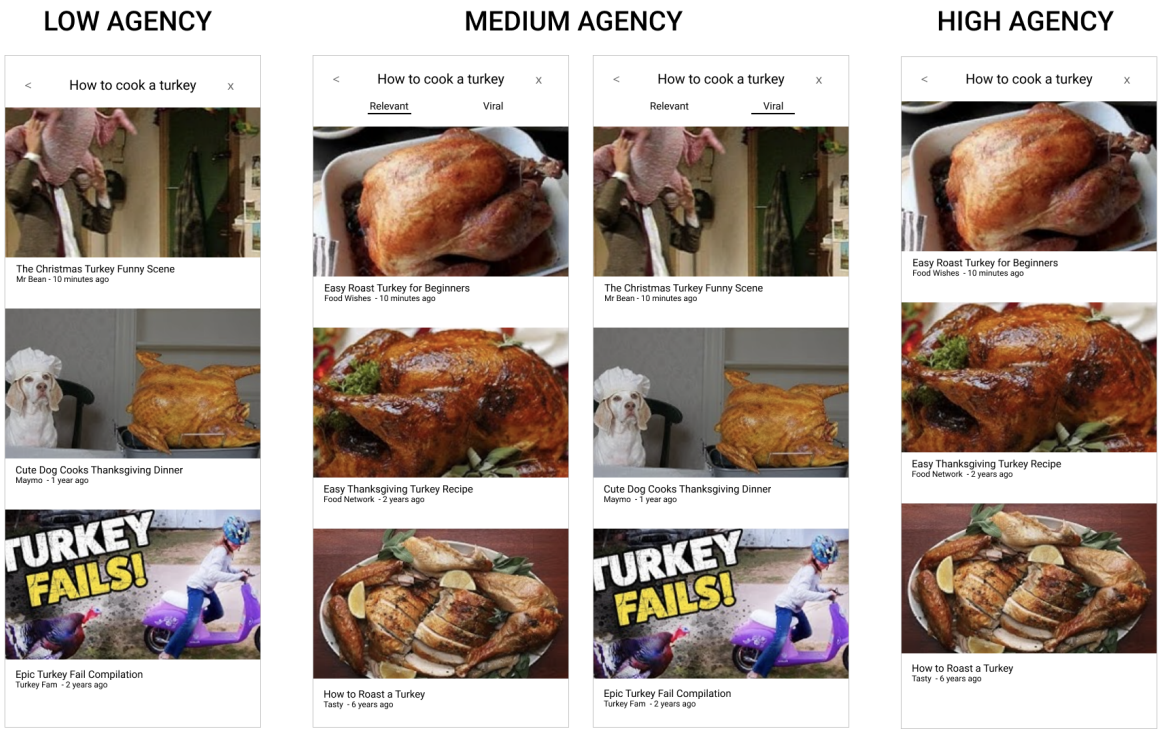


Figure 18. For the Content Selection dimension, we showed three different versions of search results that varied entertainment level to support a low, medium, or high sense of agency. Here, when a user searches for "How to cook a turkey" they see either what we called viral results (e.g., "Cute Dog Cooks Thanksgiving Dinner") or relevant results (e.g., "Easy Roast Turkey for Beginners").

We created an interactive mockup of our complete SwitchTube design in Figma and conducted usability testing with 4 participants, university students who were all 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 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 to make it easy for participants to recall the two

versions in the exit survey and interview and did not inherently suggest to participants that they should prefer one version over the other.

Table 17 shows an overview of the three different versions of the final SwitchTube study app and its features. A video overview of the app is also available here: https://youtu.be/4Np6PqeV8e8 Each of the three versions are intended to provide the user with different levels of sense of agency:

- **Explore**: Recommendations-first. The homepage shows abundant recommendations. YouTube's Data API requires the user to navigate an arduous approval process in order to permit third-party access to personalized video recommendations, so we decided to populate recommendations with top charts from several categories for the U.S. (e.g., music, comedy, and education). In other words this version was less likely to show videos that were "too personally relevant" (but not task relevant), as compared to the official YouTube app. Search results are manipulated to be more viral (by appending "viral" to the search query input by the user). The video player shows related videos below the current video and autoplays the next video. Explore was intended to provide a low level of sense of agency.
- **Focus**: Search-first. On the homepage, recommendations are turned off by default, but can be turned on by the user. Search results are relevant, just the standard YouTube results. The video player does not show related videos or autoplay. Focus was intended to provide a medium level of sense of agency.
- **Switch**: This version displays a persistent toggle at the bottom of the screen that lets the user switch between Explore Mode and Focus Mode. Switch was intended to provide a high level of sense of agency (see hypotheses in the next section).

We refer to the three versions of the app as the Explore (version), Focus (version), and Switch (version), and the two toggle options *within* Switch as Explore Mode and Focus Mode.

Table 17. Three Versions of SwitchTube

Version	Intended level of sense of agency	Homepage	Search Results	Video Player
Explore	Low	Unlimited recommendations (from 8 U.S. YouTube categories, e.g., music)	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.		

Building. An illustrated software architecture model for the SwitchTube study app on Android is shown in **Figure 19**. 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.

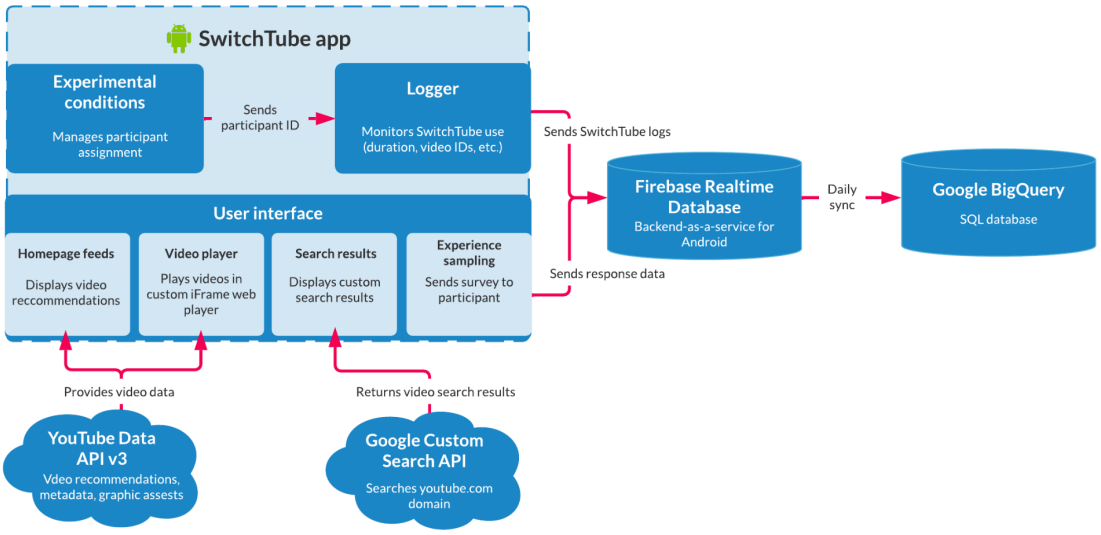


Figure 19. The technical implementation of the SwitchTube study app for Android.

One particular challenge was that we had to make due with a severely limited quota for the YouTube Data API, which we used to populate the video recommendations on the homepage and video player (related videos and autoplay) and return search results. At the time when we built SwitchTube, YouTube restricted developers to a default quota of 10,000 per day, whereas the API had previously had a far higher quota when members of our research team had previously used it [95]. As a result, we quickly maxed out our quota in our testing of the app (e.g., a search has a quota cost of 100). We tried the official form for requesting an increased quota—which we were willing to pay for—and have still not yet received a response as of a year later. 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. We hope our report of this barrier lends support to the regulatory push to require technology companies to provide researchers with greater access to audit and redesign their algorithmic systems.

Piloting. Our research team internally piloted the app on a variety of Android devices over multiple 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 of the research team for external piloting, which identified still further issues about study procedures, but also confirmed that the SwitchTube study 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 version that would be acceptable enough for participants that they would engage with it sufficiently to answer our research questions.

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 [132]. In the field of behavioral economics and social psychology, the failure to reproduce the results of many well-cited studies [177] has led to widespread adoption of pre-registration as a community norm [166]. However, it has only recently become a topic of discussion within the HCI community and is still not widely practiced [80,107,241,242]. Pre-registration can help researchers think through study protocols and guard against the natural temptation to hypothesize after the results are known (HARK) [45]. By pre-registering, we committed to testing and reporting the results of these hypotheses *regardless* of the result.

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 generally planned to use descriptive statistics rather than frequentist statistics for this purpose.

In general, our pre-registered hypotheses tested whether the Switch Version (an adaptable commitment interface) indeed 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 experience sampling responses for these metrics.

H1: User Sense of Agency. Our first set of hypotheses (H1a-H1c) addressed user **sense of agency**, which my prior work suggests is at the center of user concerns with social media and also matters for wellbeing in its own right. 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 [1]. As Switch lets users toggle between the Focus and Explore interface, we expected that this additional option will 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 is: Switch > Explore > Focus, again corresponding to 3 pairwise hypotheses that follow the same pattern as H1. In our previous YouTube work, users reported that recommendations often provided short-term satisfaction, but Focus makes these difficult to follow. We expected Switch might provide a useful option to hide 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 is: Switch > Focus > Explore, which again implies 3 pairwise comparisons. This is because users often found that search supports their personal goals, but Explore minimizes and limits search. We expect Switch to provide a useful option for times when

search does not support personal goals in comparison to recommendations (e.g., as in when survey participants said their goal was to find new or diverse content to watch).

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

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 [18], where the median is 18 participants and the mean is 53 [31]. 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 18**). We oversampled Black, Asian, and young people relative to the 2019 estimates of the United States Census Bureau [236].

Table 18. 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%), Black (20%), Asian (26%), 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. Upon multiplying these two responses together for each participant, we found that field experiment participants spent a median of 87 minutes per day in the YouTube mobile app. This is higher than the median of 70 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 users.

Procedures. As shown in **Table 19**, 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.

Table 19. 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

Upon installing SwitchTube, participants were assigned to start in either Explore or Focus following a counterbalanced assignment. Although it risked introducing ordering effects (which we address in our analysis), 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-date.

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 20**. 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.

Table 20. 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

In terms of timing, a pop-up 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 pop-up 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 [219]. 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 they 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.

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.

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 researchers (Karina Shirokova, Raveena Rao and I) who conducted a codebook thematic analysis that addressed our research questions [26]. 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 where participants described a use case of toggling between Explore and Focus during a single visit:

> *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 for collaborative affinity diagramming, where they could be easily viewed and 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 for RQ2 (switching behavior) and RQ3 (sense of agency, satisfaction, and goal alignment), which form the basis of our results.

7.4 — Results and Analysis

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

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 20**). Our analysis of usage focuses primarily on the comparison between the Explore and Focus, before turning to when and why participants used the toggle in Switch (RQ2) and how each version influenced user experience (RQ3).

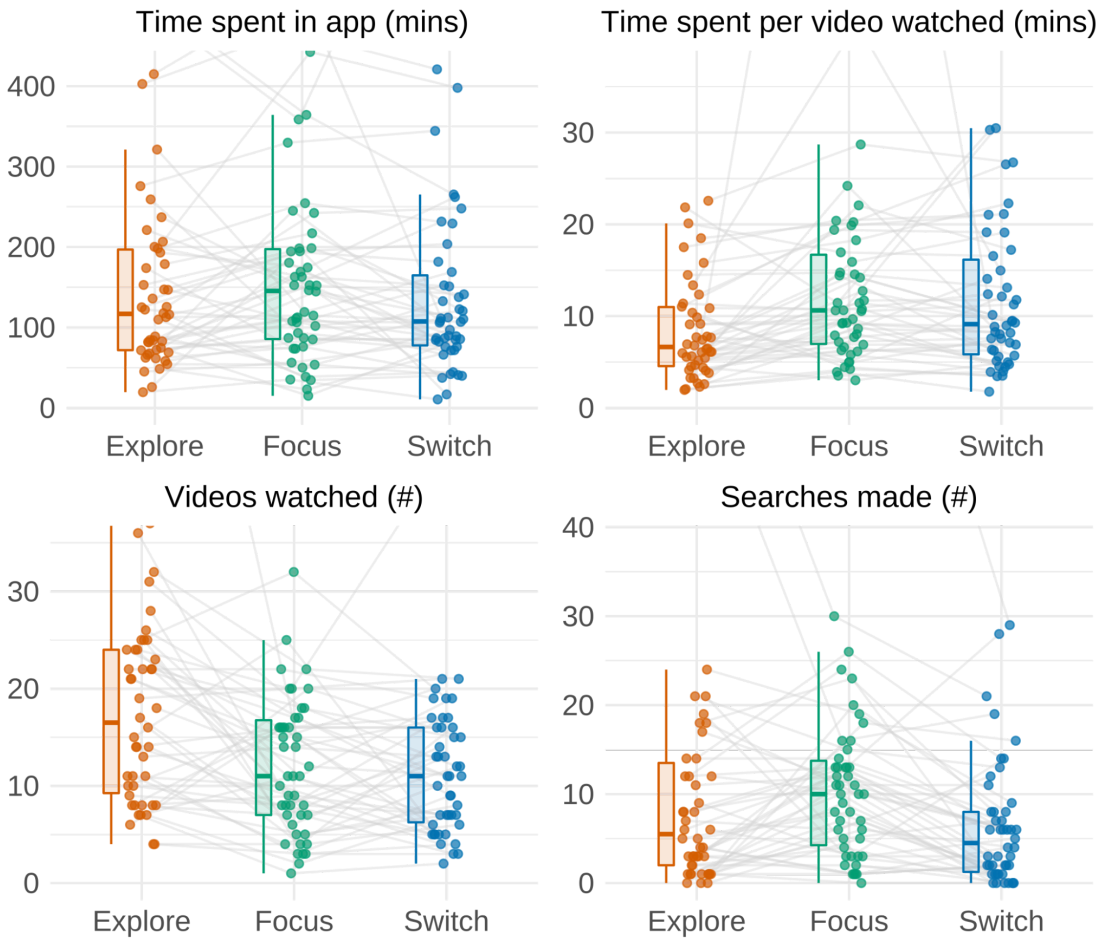


Figure 20. Log data from our deployment study with 46 participants for the three versions of the SwitchTube app: Explore, Focus, and Switch. Participants spent more time in Focus than Explore. Participants also spent more time on each video that they watched (or started) in Focus, although they actually 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]

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

[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 cut off 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.

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 entertaining 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% more of their time in SwitchTube browsing than they did in Focus and 11% 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 21**). *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 indicating that participants used features from both versions of the app, rather than just keeping the toggle on one mode.

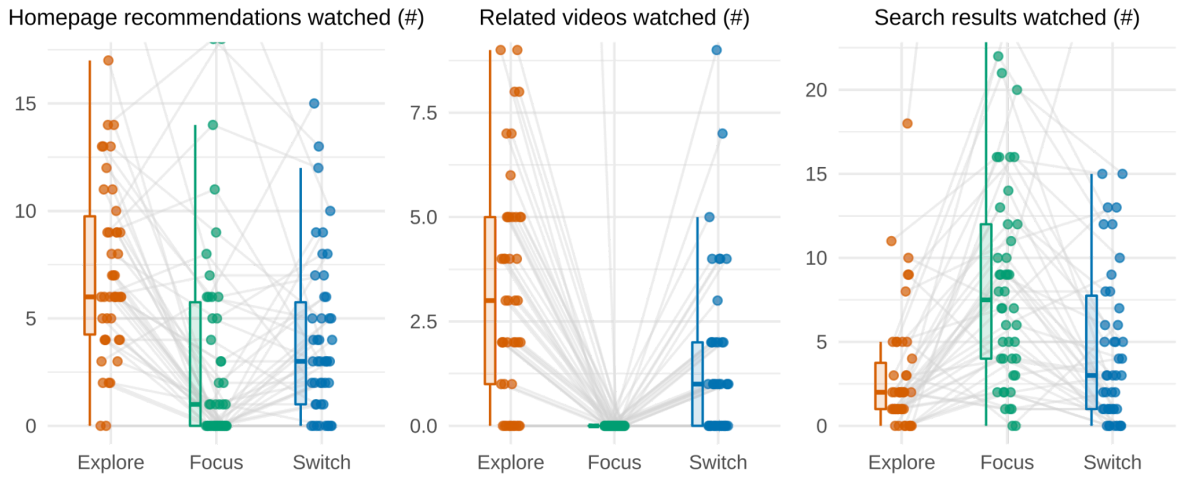


Figure 21. Videos watched from three different sources: homepage recommendations, related videos (which appeared underneath the video that was currently playing), or from searches. Video watching was driven by recommendations in Explore and by searches in Focus. In Switch, there was a balance between the two.

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.

Why and When People Switched (RQ2)

In Switch, users could toggle between the Explore and Focus Modes. 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 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 Mode 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)*

With Switch, the user could manually adapt the interface to suit their needs for that particular session.

In this use case, the adaptable interface of Switch was effective at holding the user to their commitment. For one participant, Focus was helpful and 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 and motivation for the design of SwitchTube: 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.

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 22**.

> *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 seed the system with their initial intention without getting distracted, and then explore from there. With an understanding of app usage data and switching behaviors, I now turn to how SwitchTube influenced user experience measures.

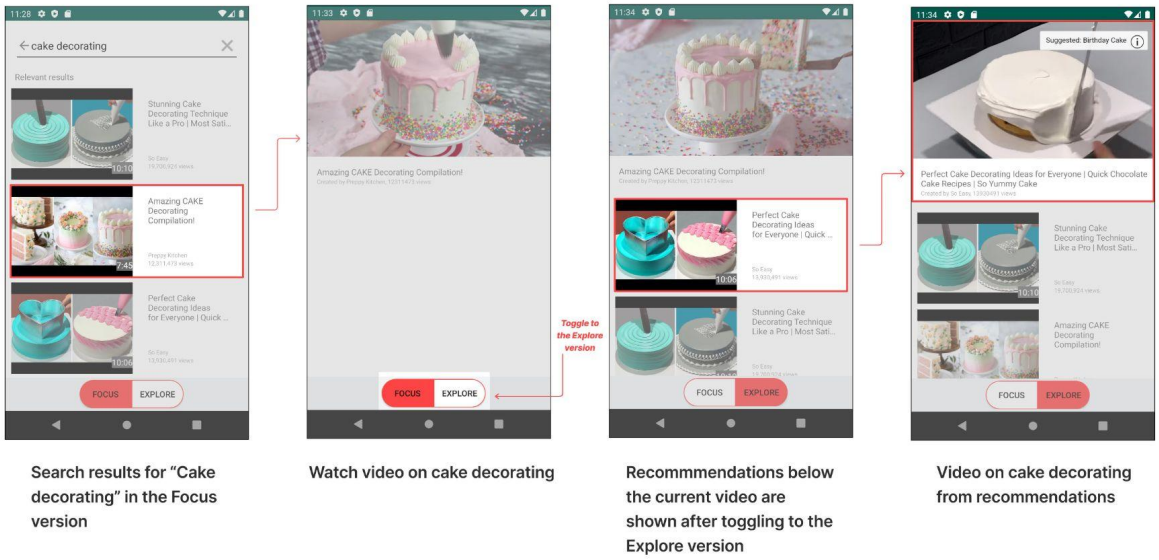


Figure 22. In this example of Use Case 4 based on an actual participant quote, the participant started with a search in Focus and then toggled to Explore to see related videos. By switching within a single session of use like this, participants were able to follow their initial intention in related directions without getting entirely sidetracked.

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 the means of each participant ratings are normally distributed or not 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. Specifically, we use the median to summarize participant means, the Friedman test for differences across all groups, and the Wilcoxon signed rank test for post-hoc pairwise comparisons.

For sense of agency, our hypothesis was: Switch > Focus > Explore. The median of the participant means was 4.1 for Explore, 5 for Focus, and 5.4 for Switch (**Figure 23**). The

participant means for control were statistically significantly different across conditions using the Friedman test of differences among repeated measures, $\chi^2(2) = 15.3$, $p<.001$. 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 ($Z=4.04$, $p<.001$) and Explore and Focus ($Z=2.71$, $p=.014$), but not between Focus and Switch ($Z=1.9$, $p=.059$). All subsequent ESM analyses use these same tests.

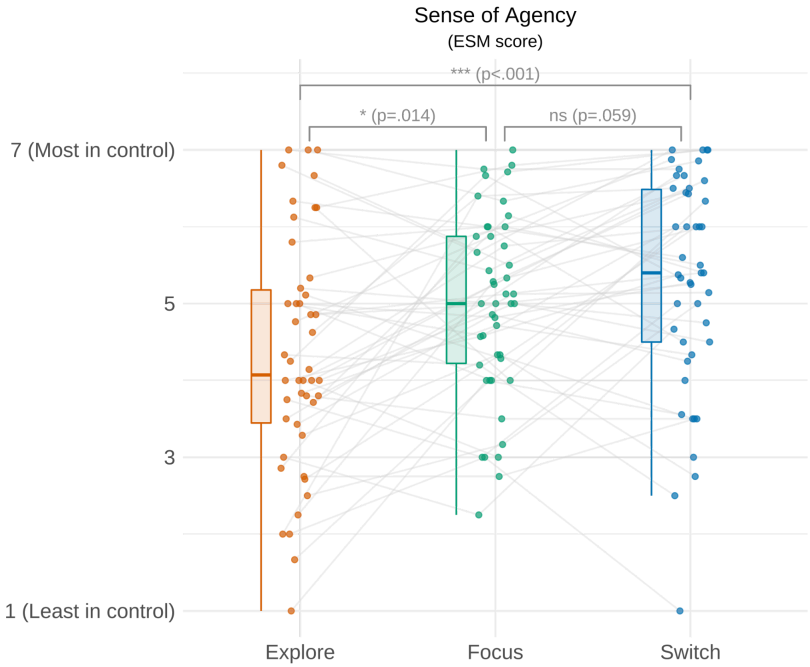


Figure 23. Experience sampling ratings for sense of agency were higher for Switch and Focus than Explore. Explore also had the greatest variation, with more participants feeling out of control as compared to the other versions. Note: Each dot represents the average of all of a single participant's ESM responses.

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.

Recommendations, and a lack of customizability thereof, 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:

> *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)*

While there was *zero* customizability for recommendations in Explore, 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") to be wholly inadequate in terms of providing the level of control they desired.

In terms of Focus versus Switch, the additional option of toggling over to Explore Mode within Switch did not increase sense of agency beyond the Focus version, but also did not diminish it. 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 one such time:

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

In other words, on occasion participants would satisfice by settling for 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).

H2: Satisfaction

For satisfaction, our hypothesis was: Switch > Explore > Focus. The median of mean scores per participant was 4.2 for Explore, 4.8 for Focus, and 5.3 for Switch (**Figure 24**.). 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 expectation, Focus was actually rated higher than Explore. Whereas we had expected that recommendations would be essential to short-term satisfaction, participants reported otherwise.

Figure 24. 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.

H3: Goal Alignment

For H3, our expectation was: Switch > Focus > Explore. The median of mean scores per participant was 4.3 for Explore, 4.8 for Focus, and 5.3 for Switch (**Figure 25**). 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 ($Z=1.73$, $p=.085$).

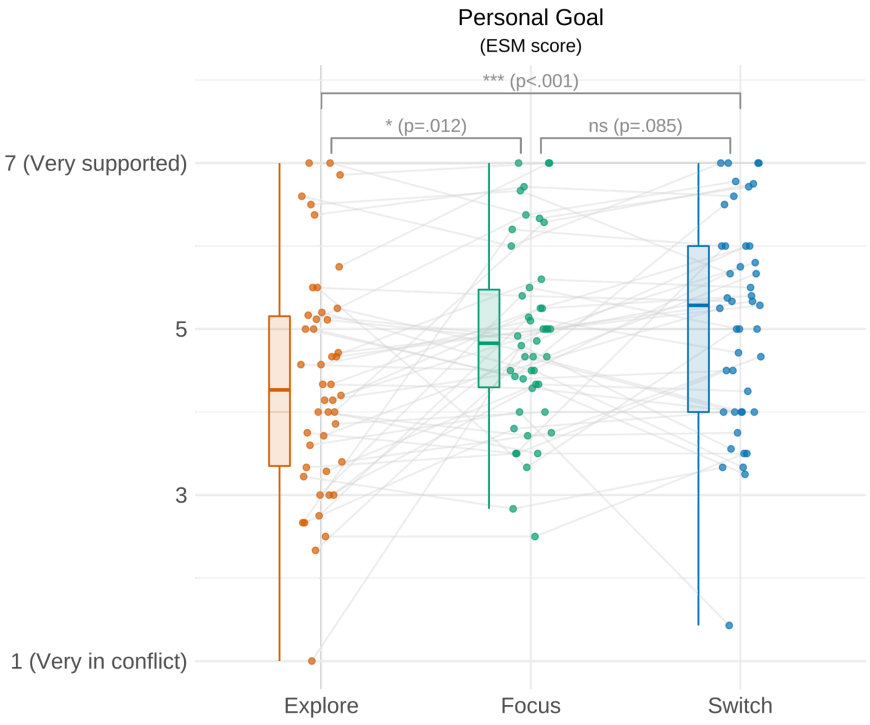


Figure 25. Experience sampling ratings for personal goal alignment were higher for Switch and Focus than Explore. A notable cluster of Switch and Focus users felt that their goals were very supported by the interface.

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, time on Focus did not induce the same feelings of guilt about ‘time-wasting.’

7.4 — 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 for 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-switching session 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**). We discuss these findings and their design implications.

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 [143], 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 [39,60,138]. Specifically, in the case of YouTube, we find that decreasing overall 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 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 [248]. 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 screen time as the singular metric of success and limited the solutions imagined by design researchers and practitioners.

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 an original goal. Unlike the concept of the goal reminders that are also used in the digital wellbeing space [94,145], nor does an ACI 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 are skilled at avoiding temptations altogether, but not necessarily at resisting them using willpower once encountered [62]. 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. Early self-regulation interventions are generally more effective [63]. Of course, this could also be done 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 a restriction of autonomy [134]. An advantage of ACIs is that they impose no such autonomy restriction: in Switch, if the user wanted to toggle over to Explore, 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 provides entertainment, but it is also a service for much more—for example, education both formal and informal, as in the SwitchTube participants who described watching videos to learn Spanish and how to fix a flat tire. Rather than offering Focus Mode only in "productivity" tools like Microsoft Word (that arguably don't need it to begin with), there's a strong case for adding it to services that people rely upon for many different use cases, such as YouTube, Twitter, and Facebook. Designers who need to convince other interested parties that a Focus Mode will not necessarily harm business interests, might cite our result that the Focus version did not reduce time spent in SwitchTube as one piece of evidence that this could potentially be a win-win solution for users *and* advertising-supported services.

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 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 [78]. Both our own findings on autoplay and previous work suggest that such a high level of automation would reduce the sense of agency [17], but it may still be the preferred interface in some

situations. Conversely, if YouTube has high confidence that the user is visiting with an intention, it could present a search-only interface and hide all recommendations.

Already, machine-learning systems can predict the specificity of a user's intent for visiting a site with some accuracy [43,96], 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 a 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 user search queries are a strong signal of the user's 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" would keep the user in a Focus Mode whereas a query for a comedian such as "Amy Schumer" would automatically switch them to Explore Mode. Such a search-informed adaptive commitment interface could offer semi-automatic support for a user's goals, but still easily afford flexibility to explore. The user provides the general direction of their intention and the system supports them to optimize for that intent, adapting into an interface for either focus or exploration.

Conclusion

Our work steps beyond the current paradigm of screen time tools and explores how a digital wellbeing design might also support users in shifting the quality of their consumption. We designed, developed, and evaluated SwitchTube, an Android app with three different versions for watching YouTube videos, in a field experiment with 46 U.S. YouTube users over three weeks. Surprisingly, Focus was rated higher than Explore in terms of user experience even though 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), in that it provided the user with the flexibility to switch, but still supported them to follow their initial commitment. We discuss implications for digital wellbeing tools and both adaptable and adaptive interfaces.

CHAPTER 8 — REFLECTING ON SENSE OF AGENCY AS A NEW LENS FOR DIGITAL WELLBEING

My work shows that the notion of screen time is insufficient for understanding and designing for digital wellbeing. Although it is easy to measure and frequently used as *the* measure of success [116,117,138], it is often a poor proxy for what users actually want. Instead, I found that smartphone users want more meaningful experiences, to change the quality of their consumption, and above all, reclaim a sense of control over their app use.

As an alternative, I advance user sense of agency over time spent as a new lens for digital wellbeing. In this discussion chapter, I examine why, when, and how we might elevate user sense of agency as a design goal. I first consider why this lens of sense of agency is sorely needed in design before turning to challenges of false sense of agency, measurement, and time-inconsistent preferences.

8.1 — The Need for Sense of Agency

Sense of agency matters for digital wellbeing in at least three ways. First, supporting user control is a common principle in HCI design guidelines [53,171]. Designing for an internal locus of control is one of Shneiderman and Plaisant's Eight Golden Rules of Interface Design, arising from the observation that users want "the sense that they are in charge of an interface and that the interface responds to their actions" [222]. As computing systems, powered by artificial intelligence, address increasingly larger spheres of human activity such as recommending the majority of the media that people consume, ensuring that this sense of agency is maintained becomes increasingly important. One of the Microsoft guidelines for Human-AI Interaction [207] is particularly apt: "Scope services when in doubt. Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals." In the case of my YouTube work, this principle could apply to video recommendations: hold off on the recommendations until the user's goal is clear, else the system risks undermining their intention.

Second, a low sense of agency over technology use predicts greater negative life effects, e.g., internet use leading to missed social activities [35]. Scales of problematic technology use generally measure both (a) lack of control and (b) negative life impacts, suggesting that 'the problem' is a combination of these two factors [38,42]. A closely related construct, self-efficacy, an individual's estimation of their capability to succeed in a future behavior, is also central to behavior change [7]. I saw a lack of self-efficacy in the users who had resigned themselves to YouTube watching habits they disliked because they felt they simply had no control over the interface. Without a sense of agency over technology use in-the-moment, it is surely hard for users to possess self-efficacy with regards to their future technology use and change their behavior as desired.

Third, and perhaps most importantly, sense of agency matters in its own right. Feeling in control of one's actions is integral to autonomy, one of the three universal basic human needs in self-determination theory [204]. More specific to technology use, a feeling of lack of control is also central to user dissatisfaction with smartphones [56,89] and Facebook use [15,42,147]. In alignment with this view, my own research on meaningfulness and the use of Facebook, Twitter, and YouTube has found that above and beyond specific concerns with technology use (e.g., a loss of sleep), users often hold a generalized concern with a loss of control, making sense of agency an apt lens through which to view user frustrations with their digital wellbeing.

8.2 — The Challenge of False Sense of Agency

A fundamental challenge to sense of agency as a measure of digital wellbeing is that people can be mistaken about their agency. Self-reported agency can be quite different from the facts of agency [164]. For example, many people continue to press 'placebo buttons' like the 'close door button' in their apartment's elevator, even when doing so has no effect [183]. In an especially pernicious example, Natasha Dow Schüll documents how slot machines in Las Vegas used to be mechanical wheels, e.g., three wheels with symbols like cherries and bells [214]. But now they are entirely decoupled from the mechanics, so they can be virtual mapped. And while consumers often still expect them to use random

mechanics, these machines are actually rigged to return "near misses." In other words, users feel a higher level sense of agency over the outcome of pulling the lever than is actually the case. But as the psychologist B. F. Skinner notes, "Almost hitting the jackpot increases the probability that the individual will play the machine, although this reinforcer costs the owner of the device nothing" [214:74]. In other words, if sense of agency catches on as a measure of digital wellbeing, designers may just dupe users into an illusion of agency.

This is a valid concern that users, designers, and regulators should guard against. I discuss approaches for a systemic and critical response to attention capture dark patterns in the next chapter. However, I also consider two factors that mitigate this concern.

First, in my estimation, such a false sense of agency over time spent is more the exception than the rule. It was far more often the case that YouTube users felt a lack of control that aligned with their actual lack of control, e.g., over the frequency, quantity, and quality of recommendations that influenced how they spent their time. Second, sense of agency need not be the only measure of success. In our SwitchTube study, we also evaluated satisfaction and goal alignment, and participants had a similar preference order for the three versions for all of these measures. In particular, measuring sense of agency in combination with whether or not a system supports a user's long-term goals should provide a strong measure of digital wellbeing. We certainly do not want users to be tricked into a sense of agency when they have little actual control of their actions and their outcomes. But nor would we want to strap users into a robot suit with artificial intelligence that consistently leads them to achieve their behavioral goals, but which affords them no sense of agency. Some combination of sense of agency *and* actual goal achievement should comprise a robust measure of digital wellbeing.

8.3 — The Challenge of Measurement

In measurement, a first challenge with sense of agency, and related constructs such as autonomy, is that there is no consensus definition [32]. Given this lack of agreement, it is also clear that measures of sense of agency will differ. In my work, I have focused on

judgements of agency, the subjective attribution of agency to the self or other (Synofzik et al. 2008). A second challenge, common in HCI, is that I wanted to assess a personality state measure that changes over time in response to interface design (using experience sampling), not a personality trait that is relatively stable over time. To complicate matters further, the Open Handbook for ESM says, “There is currently no consensus on how even very commonly investigated constructs should be measured in ESM studies” [167:89]. In this section, I discuss these challenges as well as future opportunities for measuring sense of agency.

To understand how to ask people about sense of agency, I began with an open-ended survey question to elicit qualitative responses: “What are 3 things about the mobile app that lead you to feel MOST/LEAST in control over how you spend your time on YouTube?” In piloting with 10 participants, we also tried asking about “agency,” but found that this terminology was confusing for many participants. So instead we asked about feeling "in control," which is also how sense of agency has been measured in previous studies of sense of agency in HCI (e.g., Metcalfe and Greene 2007) and on a personality trait measure (Tapal et al. 2017). This qualitative approach allowed us to ensure (a) face validity — the question was clear for lay participants; and (b) domain applicability — it yielded useful results in linking the construct of interest to our domain of interest (technology design).

Next, I sought to assess sense of agency via a personality state measure, i.e., “an ultra-brief scale that can be used to intensely and repeatedly assess states in everyday life, imposing as little burden on the participant as possible” [99:1040]. As is often the case in experience sampling, I was unable to find a suitable pre-existing measure, so I followed the most common practice: adapting a trait measure for use as a state measure [99]. Specifically, I turned again to Tapal et. al’s Sense of Agency Scale, which includes items such as “I am in full control of what I do” and “My actions just happen without my intention” [Tapal et al. 2017]. For our SwitchTube study and another study that colleagues and I conducted on a redesigned version of Twitter [248], we needed to adapt this measure so that it captured the momentary experience of engagement with our study app.

This led us to our final experience sampling measure, “For this SwitchTube use, how much did you feel out of or in control?” (1=very out of control, 7=very in control). In piloting using a talk-aloud protocol with several participants, we found that this wording was easy for participants to understand, captured judgements of agency, and varied in response to different user interfaces.

An opportunity for future research is to more systematically test the validity and reliability of state measures for sense of agency and other constructs. At least two types of validity might be considered: **First**, convergent validity, the extent to which responses to an item exhibit a strong relationship with responses on conceptually similar items. For instance, one might adapt several of Tapal et. al’s items for use in experience sampling to check whether they yield similar results. **Second**, discriminant validity, the degree to which a measure diverges from another measure whose underlying construct is conceptually unrelated. In hindsight, I wish I had of assessed discriminant validity for our three measures in the SwitchTube study (sense of agency, satisfaction, and goal alignment). Reliability could be assessed by (a) asking the same question for the same interface under similar circumstances and seeing whether responses are similar; and (b) treating each measurement occasion (e.g., each day of the study) as a separate study and assessing test-retest reliability [99]. Establishing validated measures for researchers in HCI and other fields to use would be a valuable contribution towards future research. The recently launched ESM item repository [167] is one such initiative: https://osf.io/kg376/wiki/home/

8.4 — The Challenge of Time-Inconsistent Preferences

As discussed in the SwitchTube chapter, the premise of commitment devices is to take the preferences of the present self at Time A, delegate them to a commitment device (or interface), which then enforces them upon the future self at Time B. Implied is that an individual’s preferences are inconsistent over time, otherwise there would be no need to enforce anything upon the future self in the first place. And in the case of digital device use

this is indeed often the case (e.g., I might decide to watch another recommended video on YouTube instead of going to bed, only to regret it later [44]).

This challenge can be viewed through different conceptual lenses. The psychologist Daniel Kahneman highlights a gap between the experiencing and the remembering self [104]. A person may listen to a classical music concert in a state of rapt bliss for an hour, but say that 'the experience was ruined' by a screeching violin in the final minute. In this case, Kahneman points out that it is not the experience that was ruined, but rather the *memory* of the experience. At first, Kahneman was drawn to the neat conclusion that life is simply the sum of moment-to-moment pleasure and that the account of the remembering self should be seen as a cognitive error. However, his experience sampling research found that life is full of experiences that are often miserable in the moment, but deeply satisfying upon reflection (e.g., raising children), leading him to call for balance when prioritizing between the experiencing and the remembering selves.

In my own early work, I explored the concept of meaningfulness, based on the eudaimonic tradition of living life with a sense of fulfillment and purpose, as a way of understanding the gap between people's intentions and their actions in the case of smartphone use [139]. Similarly, Mekler and colleagues propose designing for enduring meaning rather than just the sum of momentary pleasure over time [36,156,157]. Alternatively, Baumer et al. propose sense of agency as an organizing concept for social media use, which considers whether users are aware of their actions, feel capacity to change their actions, and view their actions as intentional, in the sense of being aligned the goals of their reflective self [15]. It is this concept of sense of agency that I draw upon throughout my work on YouTube.

In terms of *when* to measure sense of agency in the face of time-inconsistent preferences, my primary approach has been to use ESM to prompt for an *in situ* response from the reflective self at many different time points. That is, in both my meaningfulness and SwitchTube studies, our research team instrumented our study app to prompt participants to (a) respond to our state measures immediately *after* using the app of

interest for a while (e.g., 30 seconds minimum in the SwitchTube study) and (b) to do so in a way that prompts the participant to engage in what we might call 'micro-reflection': "For this SwitchTube use, how much did you feel out of or in control?" (1=very out of control, 7=very in control). This has allowed me to capture the user in different states, in some of which the participant's long-term goals may be more or less salient. This has also led to a recognition that a high sense of agency is not desired all of the time, but rather that there are also times when the user just wants to leanback and let the system take control. Designers should support both use cases.

An interesting question here is whether the responses from these short-form responses align with the long-form ones might elicit in a survey or interview. Since both of these studies have *also* included an exit survey and interview with a subset of participants, I can report that my analysis has generally found a strong alignment between an individual's overall experience sampling responses and their survey and interview reflections. However, by using data-driven retrospective interviews, in which we showed participants their experience sampling data when we asked them to reflect, participants may have provided different responses than if they had relied solely on recall. At the same time, a powerful benefit of this approach was that it tied participant responses to specific design features and a specific context, the kind of contextually rich findings that help inform work in HCI.

Conclusion

My work demonstrates that sense of agency is a valuable lens for understanding digital wellbeing that can help designers move beyond the screen time paradigm. Yet challenges remain: users can sometimes feel a false sense of agency and there are still open questions with regards to both how and when to measure sense of agency. As an overall approach, I recommend triangulation. That is rather than relying upon sense of agency as the sole measure of success, also consider whether consider whether a design brings users closer to their goals and whether similar results are obtained if sense of agency is measured using slightly different wording and at different timings, e.g., in both

experience sampling and exit interviews. This approach of triangulating data from multiple sources of data can yield rich results that tie the subjective user experience to the specific design features within a technology.

CHAPTER 9 — FUTURE WORK AND IMPACT: INFRASTRUCTURAL, SOCIAL, AND SYSTEMIC CONSIDERATIONS

My work also suggests opportunities for future work. **First**, in the process of conducting my research it took considerable time and effort to get an overview of the mobile app ecosystem. How might we develop tools and best practices for conducting systematic app reviews? **Second**, my research has focused on sense of agency for the individual user, but how might this concept apply to social groups, such as families with tweens? **Third**, rathering than burdening individuals or relying on companies to change, how might we take systemic approaches to supporting sense of agency over technology use?

9.1 — Research Infrastructure for Systematic App Reviews

To address digital wellbeing and other issues with digital technologies, it is important to have a systematic understanding of the mobile app ecosystem. In my own work, I have conducted systematic app reviews (SARs), in which I systematically collected and analyzed data from mobile app stores (e.g., Google Play) in two domains: digital self-control tools [143] and mindfulness apps [135]. However, I found that whereas well-established best practices exist for systematic literature reviews (e.g., PRISMA reporting standards [162]), best practices for systematic app reviews are unclear and the process of data collection from mobile app stores is costly and challenging. For example, how different are results across Google Play and iOS and when do researchers need to review apps from both? Each SAR entails dozens of such methodological questions for which there is currently little guidance on the considerations and tradeoffs. How might we develop best practices and tools that make it easy for non-technical researchers across domains to conduct SARs? I plan to address this challenge in my own future work.

I conceptualize SARs as the process of collecting and analyzing data from mobile app stores (e.g., Google Play) to gain a systematic understanding of the state of the mobile

app ecosystem and/or the features and practices *within* apps in a domain such as digital wellbeing, mental health, or privacy. For example, a SAR by psychology researchers might examine adherence to clinical guidelines in apps that contain the keyword "cognitive behavioral therapy" in the description and have been installed over a million times. A SAR by privacy researchers might use an automated analysis of application packages to check for the prevalence of third-party trackers *within* children's apps, where targeted advertising is illegal in the U.S. and EU. SARs might provide an empirical understanding of the app ecosystem in a specific domain, inform regulatory policy, and/or inspire ideas for new mobile apps.

To support SARs, there are at least three challenges that researchers in the field of human-computer interaction might address.

- **First**, researchers could systematically review the literature to understand existing goals and practices for SARs, both in a field with non-technical researchers (e.g., health) and one with technical researchers (e.g., HCI). The former should be more likely to identify technical barriers that researchers face while the latter should be able to identify technical possibilities for this method (e.g., techniques for the automated analysis of the content within applications).

- **Second**, researchers could conduct user research on how other researchers conduct SARs. One approach could be to observe research teams as they conduct SARs. Another approach could be to go beyond observing and provide technical consulting for researchers from different domains who wish to conduct SARs. Both of these approaches could contribute an understanding of user needs, while the latter could move towards prototyping scalable solutions.

- **Third,** researchers could leverage knowledge of researcher needs to develop best practices and self-service open source tools that enable any researcher to conduct a SAR on their own. Best practices and considerations might be proposed for each step of the SARs process, e.g., scoping → data collection → analysis → reporting. As a first step, self-service tools might include an open source scraper for

downloading metadata from app stores with an easy-to-use graphical user interface (rather than just an API that current tools use, e.g., [256]). As a next step, this might include tools that let researchers analyze the content within mobile app packages themselves).

Developing this SARs research infrastructure should support researchers across domains to develop a richer understanding of the mobile app ecosystem.

9.2 — Social Considerations

My work so far has focused on app design and the individual user. However, technology and social media use in particular are highly influenced by social context. In this section, I discuss opportunities for future work that looks beyond the individual user and investigates how social groups can collaboratively achieve a sense of agency over phone use. I focus on how families with children manage phone use as an important and challenging domain that I intend to research in the near future.

Parents and teens both sometimes struggle to align the time they spend on phones and tablets with their personal values, and also often disagree with each other about what constitutes desired use. Popular parental control apps enable parents to monitor and restrict their child's device use and 57% of parents report setting screen time restrictions on their teen [185]. Yet these tools largely neglect child values for their *own* use and frustrations with their parents' use [97]. Ghosh et al. conducted a SAR of parental control apps and found that children feel that existing apps are "overly restrictive and invasive of their personal privacy, negatively impacting their relationships with their parents" [75]. This one-sided tool design denies children and teens a sense of agency over their own technology use and practice in self-regulation.

One step forward could be to investigate child needs and re-envision smartphone use management interventions from a family-centered perspective. Ko et al. developed, deployed, and evaluated FamiLync, an app for limiting smartphone use that takes a parental mediation approach wherein the screen time of all family members is displayed

on a family dashboard [112]. They find that FamiLync reduced screen time and improved family awareness. What family-centered approaches might similarly support teen sense of agency over technology use while still meeting parental goals for their child's use?

9.3 — Systemic Considerations

I am often asked: "How will your research have an impact given the business incentives of the attention economy?" To a certain extent, I believe that my work can still have an impact at the individual and corporate level. Yet the largest opportunity for change is at the systemic level. Here I discuss these opportunities for change and how they might be advanced in future work.

For the individual user, my research has identified and demonstrated the efficacy of publicly-available tools and strategies that they might adopt to support their sense of agency. For instance, newsfeed blockers or goal reminders can help users manage how they spend their time on Facebook [145]. More generally, I have examined hundreds of publicly-available tools on app stores and web browser stores that offer users some ability to architect their digital environment so that it supports their needs [143].

Yet placing the entire or primary burden of managing technology use on the individual user in the face of apps that are designed for attention capture and which individuals often feel obligated to use for social or employment purposes is both unmanageable and unfair. As Tristan Harris, co-founder of 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" [10]. The term "cruel optimism," coined by Lauren Berlant [19], also describes the situation when users are told: "Some people manage their technology use just fine, why can't you just do the same?" [88] Just as with diet, fitness, and addictive behaviors, such advice is often not only unhelpful in that it does not work for the vast majority of people, but is also demoralizing and cruel.

One of the limitations of behavioral economics is that it focuses on the individual rather than systemic factors. In behavioral economics, nudging or behavior change

interventions are grounded in the heuristics and biases of the individual. This can be powerful when the individual themself has the means, knowledge, and skills to self-nudge their own behavior or when a benevolent actor does so on their behalf. However, users rarely have such resources or allies when it comes to digital interfaces they use. And behavioral economics interventions do not typically address systemic issues of power, representation, and incentives.

For technology corporations, my work identifies design changes that support user sense of agency as validated in field experiments. For a designer at a firm, the first change I would recommend is tracking a reflective measure of user wellbeing that varies in response to specific features. In my work, I have focused on sense of agency, but my SwitchTube results suggest that tracking satisfaction or goal alignment might yield similar results. Having a metric to point to can help a designer not only test the influence of their own digital wellbeing designs, but also convince other skeptical parties that their work is improving the user experience. As discussed in the section on measurement in the previous chapter, future work could establish single-item measures with a low user burden that are acceptable in industry.

In some cases, such changes might be win-win between user and business interests. For example, our SwitchTube finding that Focus improved user experience but did not decrease time spent as compared to Explore. This suggests that, for apps with multiple use cases such as YouTube, introducing a Focus version will not necessarily reduce time spent and revenues. In other cases, it may be that designing for sense of agency decreases engagement in the short term, but will increase user loyalty in the long term. For example, in January 2018, CEO Mark Zuckerberg announced that the social network's news feed would prioritize "meaningful interactions" over passive experiences [84]. He shared that he believes this will hurt his firm's engagement measures in the short run, but make the service more valuable to users in the long run. It is worth noting that actors with longer investment time horizons, such as platform owners and apps with low customer churn rather than those with more transactional one-off interactions, are in better position to benefit from long-term user engagement.

Unfortunately Facebook's actual implementation of M.S.I. (meaningful social interactions) as a metric is also a cautionary lesson for researchers and designers. The first iteration of changes to the Facebook algorithm was rolled out in a little over one month. It largely used existing measures of engagement, e.g., likes and comments, as a proxy for meaningfulness. The problem was that the most-liked and most-commented posts which got boosted in the newsfeed tended to be the ones that provoked fear and outrage [82]. A reflective, self-reported measure of user wellbeing might provide a better signal of meaningful experience.

When companies do not act, systemic changes are called for. The even harder case is when the user would benefit from changes that harm the firm's economic interests. In particular, regulators can and should make changes to support greater user sense of agency. One promising approach is regulations at the level of design patterns, informed by the research of HCI scholars. For researchers, a language of 'attention capture dark patterns' might serve as a way of understanding and communicating the exploitative tactics that commercial technologies sometimes use to capture attention. Thaler and Sunstein define nudging as architecting choices to align with the desires of the reflective self, the same approach can and is widely used to encourage actions that are misaligned with the reflective self. Manipulation of the automatic system is ubiquitous amongst the technologies that people report are most prone to direct their attention in ways that are misaligned with their goals [139]: autoplay on Netflix, infinite scroll on Facebook, and the emphasis upon lost items when leaving the game in Farmville. Yet these designs are not yet recognized by practitioners or academics as a class of dark patterns [79], perhaps because their costs are 'merely' attentional as opposed to financial, social, or privacy-related [133]. Colleagues and I are currently working to propose criteria for attention capture dark patterns.

A pattern language can be especially effective as a method of communication when presented alongside a library of exemplars. Design patterns are easily understood by academics and practitioners alike, meaning that they could aid in the important task of translating across the research-practice gap [57:49]. As a critical perspective, dark

patterns (or dark nudges) may also hold a particular resonance with the public, media, and regulators [79]. Rather than communicating predominantly with other researchers via academic papers, establishing a pattern language and library as a public resource could spark a richer conversation with a wider community, which could also contribute back to the findings of the research itself [48].

In terms of actual regulatory action, the EU has already banned some of the most exploitative industry practices like opt-out email lists in part due to public input against dark patterns [252]. The EU is also currently circulating a draft of the Digital Services Act that would require large online platforms to [257] provide transparency into the recommender systems they use and the option of using a recommender system where content is not based on profiling. Many of the participants in the YouTube studies might benefit from these changes. Based on my research, regulators might also consider requiring these platforms to provide the option of a Focus Mode or search-based interface, rather than only a recommendations-based one. Another change that regulators might consider and researchers should investigate is to require platforms (e.g., Facebook) to provide a paid service, not only an ad-supported one. While this introduces other barriers to access, it would also provide an option to some users and force platforms to consider and build alternative modes of monetization. Addressing systemic issues through regulation can help achieve a digital environment that better supports sense of agency and digital wellbeing.

Conclusion

Sense of agency is a helpful concept for understanding and designing for digital wellbeing. As a lens for understanding, sense of agency moves away from a reductionist paradigm that digital wellbeing is merely about minimizing screen time. Instead, it challenges designers to consider which aspects of digital experiences truly support the user's intentions. Whereas current screen time tools largely rely upon external mechanisms that indiscriminately block access to desired features, a design approach grounded in sense of agency could let users adjust or "switch" the internal mechanisms of the services they use to suit their own needs. Finally, my work also reveals the extent to which sense of agency is more than just a means to an end, but also an outcome that itself matters deeply for users in their interactions with digital technologies.

References

1. Jesper Aagaard. 2015. Drawn to distraction: A qualitative study of off-task use of educational technology. *Computers & education* 87: 90–97. https://doi.org/10.1016/j.compedu.2015.03.010
2. Alexander T. Adams, Jean Costa, Malte F. Jung, and Tanzeem Choudhury. 2015. Mindless Computing: Designing Technologies to Subtly Influence Behavior. *Proceedings of the ... ACM International Conference on Ubiquitous Computing . UbiComp (Conference)* 2015: 719–730. https://doi.org/10.1145/2750858.2805843
3. Elena Agapie. 2020. Designing for Human Supported Evidence-Based Planning. digital.lib.washington.edu. Retrieved from https://digital.lib.washington.edu/researchworks/handle/1773/45709
4. Morgan G. Ames. 2013. Managing Mobile Multitasking: The Culture of iPhones on Stanford Campus. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (CSCW '13), 1487–1498. https://doi.org/10.1145/2441776.2441945
5. Apple Inc. Details for app privacy questions now available - News - Apple Developer. Retrieved September 13, 2020 from https://developer.apple.com/news/?id=hx9s63c5
6. Catherine Baab-Muguira. 2017. The Stupidly Simple Productivity Hack Hiding In Microsoft Word. Retrieved September 11, 2020 from https://www.fastcompany.com/3068825/the-stupidly-simple-productivity-hack-hiding-in-microsoft-word
7. Albert Bandura and Nancy E. Adams. 1977. Analysis of self-efficacy theory of behavioral change. *Cognitive therapy and research* 1, 4: 287–310. https://doi.org/10.1007/BF01663995
8. Liam Bannon, John Bowers, Peter Carstensen, John A. Hughes, K. Kuutti, James Pycock, Tom Rodden, Kjeld Schmidt, Dan Shapiro, Wes Sharrock, and Others. 1994. *Informing CSCW system requirements*. Lancaster University. Retrieved from https://www.forskningsdatabasen.dk/en/catalog/2185760093
9. Nikola Banovic, Christina Brant, Jennifer Mankoff, and Anind Dey. 2014. ProactiveTasks: the short of mobile device use sessions. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (MobileHCI '14), 243–252. https://doi.org/10.1145/2628363.2628380
10. A. Baron, M. Perone, and M. Galizio. 1991. Analyzing the reinforcement process at the human level: can application and behavioristic interpretation replace laboratory research? *The Behavior analyst / MABA* 14, 2: 95–105. https://doi.org/10.1007/BF03392557
11. Naomi S. Baron. 2010. *Always On: Language in an Online and Mobile World*. Oxford University Press. Retrieved from https://market.android.com/details?id=book-pq4PJv3xZ5UC
12. Roy F. Baumeister, Kathleen D. Vohs, Jennifer L. Aaker, and Emily N. Garbinsky. 2013. Some key differences between a happy life and a meaningful life. *The journal of positive psychology* 8, 6: 505–516. https://doi.org/10.1080/17439760.2013.830764
13. 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* (CHI '13), 3257–3266. https://doi.org/10.1145/2470654.2466446
14. Eric P. S. Baumer and M. Six Silberman. 2011. When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2271–2274. https://doi.org/10.1145/1978942.1979275
15. 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: 23:1–23:19. https://doi.org/10.1145/3274292
16. J. T. Behrens - Psychological Methods and 1997. 1997. Principles and procedures of exploratory data analysis. *psycnet. apa. org*. Retrieved from http://psycnet.apa.org/journals/met/2/2/131/
17. Bruno Berberian, Jean-Christophe Sarrazin, Patrick Le Blaye, and Patrick Haggard. 2012. Automation technology and sense of control: a window on human agency. *PloS one* 7, 3: e34075. https://doi.org/10.1371/journal.pone.0034075
18. Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Comput. Surv.* 50, 6: 93:1–93:40. https://doi.org/10.1145/3123988
19. Berlant. 2010. Cruel optimism. *The affect theory reader*. Retrieved from https://books.google.com/books?hl=en&lr=&id=bl0udWQii48C&oi=fnd&pg=PA93&dq=cruel+optimism&ots=nrAb9IjY1P&sig=P3C_F09HExgSPR0uUUHbVGJvqXI
20. Katherine Bessiere, Sara Kiesler, Robert Kraut, and Bonka S. Boneva. 2008. Effects of Internet use and social resources on changes in depression. *Information, Community & Society* 11, 1: 47–70.
21. Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. 2011. Falling Asleep with Angry Birds, Facebook and Kindle: A Large Scale Study on Mobile Application Usage. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (MobileHCI '11), 47–56. https://doi.org/10.1145/2037373.2037383
22. Khatereh Borhani, Brianna Beck, and Patrick Haggard. 2017. Choosing, Doing, and Controlling: Implicit Sense of Agency Over Somatosensory Events. *Psychological science* 28, 7: 882–893. https://doi.org/10.1177/0956797617697693
23. Bianca Bosker. 2016. What Will Break People's Addictions to Their Phones? *The Atlantic*. Retrieved April 30, 2018 from http://www.theatlantic.com/magazine/archive/2016/11/the-binge-breaker/501122/
24. Richard E. Boyatzis. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development*. SAGE. Retrieved from https://play.google.com/store/books/details?id=_rfClWRhIKAC
25. Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2: 77–101. https://doi.org/10.1191/1478088706qp063oa
26. Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4: 589–597. https://doi.org/10.1080/2159676X.2019.1628806
27. 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
28. Harry Brignull and Alexander Darlington. What are dark patterns? *Dark Patterns*. Retrieved September 28, 2019 from https://www.darkpatterns.org/
29. Gharad Bryan, Dean Karlan, and Scott Nelson. 2010. Commitment Devices. *Annual review of economics* 2, 1: 671–698. https://doi.org/10.1146/annurev.economics.102308.124324
30. Peter Buxmann, Hannah Krasnova, H. Wenninger, and T. Widjaja. 2013. Envy on Facebook: A Hidden Threat to Users' Life Satisfaction. In *11th International Conference on Wirtschaftsinformatik*.
31. Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 981–992. https://doi.org/10.1145/2858036.2858498

32. Ryan Calo. 2013. Digital market manipulation. *The George Washington law review* 82: 995. Retrieved from https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/gwlr82&section=34&casa_token=TD7Dj3IMGygAAAAA:j6KYmBRy_FXY1GdfpopXhHyT8tQ6SqkfeJzVSSNO9JN4rPPAcI91tTpj5kDOboGKCT-50MN6eQ
33. Rafael A. Calvo, Karthik Dinakar, Rosalind Picard, and Pattie Maes. 2016. Computing in Mental Health. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16), 3438–3445. https://doi.org/10.1145/2851581.2856463
34. Rafael A. Calvo and Dorian Peters. 2014. *Positive Computing: Technology for Wellbeing and Human Potential*. MIT Press. Retrieved from https://market.android.com/details?id=book-uI6ZBQAAQBAJ
35. Scott E. Caplan. 2010. Theory and measurement of generalized problematic Internet use: A two-step approach. *Computers in human behavior* 26, 5: 1089–1097. https://doi.org/10.1016/j.chb.2010.03.012
36. Vanessa Julia Carpenter and Elisa D. Mekler. 2019. Towards Metrics of Meaningfulness for Tech Practitioners. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts). ACM, New York, NY, USA*. Retrieved from https://www.researchgate.net/profile/Elisa_Mekler/publication/330656251_Towards_Metrics_of_Meaningfulness_for_Tech_Practitioners/links/5c4c8968299bf12be3e5779d/Towards-Metrics-of-Meaningfulness-for-Tech-Practitioners.pdf
37. Nicholas Carr. 2011. *The shallows: What the Internet is doing to our brains*. WW Norton & Company. Retrieved from https://books.google.com/books?hl=en&lr=&id=1KayoVl3OTMC&oi=fnd&pg=PA1&dq=the+shallows&ots=3G6XwPTkYL&sig=YkxY3G79GPZSyZRRUOCLHmltqec
38. Hilarie Cash, Cosette D. Rae, Ann H. Steel, and Alexander Winkler. 2012. Internet Addiction: A Brief Summary of Research and Practice. *Current psychiatry reviews* 8, 4: 292–298. https://doi.org/10.2174/157340012803520513
39. 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* (CHI EA '19), 1–8. https://doi.org/10.1145/3290607.3298998
40. Michael Chan. 2015. Mobile phones and the good life: Examining the relationships among mobile use, social capital and subjective well-being. *New Media & Society* 17, 1: 96–113. https://doi.org/10.1177/1461444813516836
41. Angela Chen. A new bill would ban making social media too addictive. *MIT Technology Review*. Retrieved August 27, 2020 from https://www.technologyreview.com/2019/07/30/133976/josh-hawley-social-media-addictive-design-legislation-smart-act-bill/
42. Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding Perceptions of Problematic Facebook Use: When People Experience Negative Life Impact and a Lack of Control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 199. https://doi.org/10.1145/3290605.3300429
43. 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*, 593–601. https://doi.org/10.1145/3041021.3054198
44. 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: 1–36.

https://doi.org/10.1145/3479600
45. 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* (CHI '18), 1–12. https://doi.org/10.1145/3173574.3173715
46. Akiba A. Cohen, Mark R. Levy, and Karen Golden. 1988. Children's Uses and Gratifications of Home VCRs: Evolution or Revolution. *Communication research* 15, 6: 772–780. https://doi.org/10.1177/009365088015006006
47. Emily I. M. Collins, Anna L. Cox, Jon Bird, and Cassie Cornish-Tresstail. 2014. Barriers to Engagement with a Personal Informatics Productivity Tool. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (OzCHI '14), 370–379. https://doi.org/10.1145/2686612.2686668
48. Lucas Colusso, Cynthia L. Bennett, Gary Hsieh, and Sean A. Munson. 2017. Translational Resources: Reducing the Gap Between Academic Research and HCI Practice. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS '17), 957–968. https://doi.org/10.1145/3064663.3064667
49. Council on Communications and Media. 2016. Media and Young Minds. *Pediatrics* 138, 5. https://doi.org/10.1542/peds.2016-2591
50. Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (RecSys '16), 191–198. https://doi.org/10.1145/2959100.2959190
51. Alex Cowan. 2014. Making Your Product a Habit: The Hook Framework. *Cowan+*. Retrieved April 16, 2022 from https://www.alexandercowan.com/the-hook-framework/
52. 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* (CHI EA '16), 1389–1397. https://doi.org/10.1145/2851581.2892410
53. David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I did that! Measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 2025–2034. https://doi.org/10.1145/2207676.2208350
54. M. Csikszentmihalyi and R. Larson. 1987. Validity and reliability of the Experience-Sampling Method. *The Journal of nervous and mental disease* 175, 9: 526–536. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3655778
55. Richard J. Davidson and Cortland J. Dahl. 2018. Outstanding Challenges in Scientific Research on Mindfulness and Meditation. *Perspectives on psychological science: a journal of the Association for Psychological Science 13*, 62–65. https://doi.org/10.1177/1745691617718358
56. Katie Davis, Anja Dinhopl, and Alexis Hiniker. 2019. " Everything's the Phone": Understanding the Phone's Supercharged Role in Parent-Teen Relationships. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 227. Retrieved from https://dl.acm.org/citation.cfm?id=3300457
57. Andy Dearden and Janet Finlay. 2006. Pattern Languages in HCI: A Critical Review. *Human–Computer Interaction* 21, 1: 49–102. https://doi.org/10.1207/s15327051hci2101_3
58. Edward L. Deci and Richard M. Ryan. 2008. Hedonia, eudaimonia, and well-being: an introduction. *Journal of happiness studies* 9, 1: 1–11. https://doi.org/10.1007/s10902-006-9018-1
59. 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: 1157–1167. https://doi.org/10.1002/bdm.2031
60. M. A. Devito, A. M. Walker, J. Birnholtz, and K. Ringland. 2019. Social Technologies for Digital

Wellbeing Among Marginalized Communities. *Publication of the 2019 ….* Retrieved from https://dl.acm.org/doi/abs/10.1145/3311957.3359442
61. Anind K. Dey, Katarzyna Wac, Denzil Ferreira, Kevin Tassini, Jin-Hyuk Hong, and Julian Ramos. 2011. Getting closer: an empirical investigation of the proximity of user to their smart phones. In *Proceedings of the 13th international conference on Ubiquitous computing* (UbiComp '11), 163–172. https://doi.org/10.1145/2030112.2030135
62. Angela L. Duckworth and James J. Gross. 2020. Behavior Change. *Organizational behavior and human decision processes* 161, Suppl: 39–49. https://doi.org/10.1016/j.obhdp.2020.09.002
63. 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: 329–341. https://doi.org/10.1037/edu0000062
64. Matthew S. Eastin, Terry Daugherty, and Neal M. Burns. 2010. *Handbook of Research on Digital Media and Advertising: User Generated Content Consumption: User Generated Content Consumption*. IGI Global. Retrieved from https://play.google.com/store/books/details?id=pY2SszHlFA4C
65. Michael D. Ekstrand and Martijn C. Willemsen. 2016. Behaviorism is Not Enough: Better Recommendations Through Listening to Users. In *Proceedings of the 10th ACM Conference on Recommender Systems* (RecSys '16), 221–224. https://doi.org/10.1145/2959100.2959179
66. Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook "friends:" social capital and college students' use of online social network sites. *Journal of computer-mediated communication: JCMC* 12, 4: 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x
67. Nir Eyal. 2014. *Hooked: How to build habit-forming products*. Penguin UK. Retrieved from http://www.weltderfertigung.de/downloads/buch-hooked.pdf
68. Robert Felner, Angela Adan, Richard Price, E. L. Cowen, R. P. Lorion, and J. Ramos-McKay. 1988. 14 Ounces of prevention: A casebook for practitioners.
69. C. B. Ferster and B. F. Skinner. 1957. Appleton-Century-Crofts, East Norwalk, CT, US. https://doi.org/10.1037/10627-000
70. B. J. Fogg. 2009. A Behavior Model for Persuasive Design. In *Proceedings of the 4th International Conference on Persuasive Technology* (Persuasive '09), 40:1–40:7. https://doi.org/10.1145/1541948.1541999
71. Donelson R. Forsyth. 2008. Self-serving bias. Retrieved from https://scholarship.richmond.edu/cgi/viewcontent.cgi?article=1164&context=jepson-faculty-publications
72. Milton Friedman and Rose Friedman. 1990. *Free to Choose: A Personal Statement*. Houghton Mifflin Harcourt. Retrieved from https://market.android.com/details?id=book-F5z1B5SwGUEC
73. Peter Batya Friedman, H. Kahn, and Alan Borning. 2020. Value Sensitive Design and Information Systems. *The Ethics of Information Technologies*, 289–313. https://doi.org/10.4324/9781003075011-21
74. Bin Fu, Jialiu Lin, Lei Li, Christos Faloutsos, Jason Hong, and Norman Sadeh. 2013. Why People Hate Your App: Making Sense of User Feedback in a Mobile App Store. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '13), 1276–1284. https://doi.org/10.1145/2487575.2488202
75. Arup Kumar Ghosh, Karla Badillo-Urquiola, Shion Guha, Joseph J. LaViola Jr, and Pamela J. Wisniewski. 2018. Safety vs. Surveillance: What Children Have to Say about Mobile Apps for Parental Control. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–14. https://doi.org/10.1145/3173574.3173698
76. Peter M. Gollwitzer. 1999. Implementation intentions: Strong effects of simple plans. *The*

American psychologist 54, 7: 493. https://doi.org/10.1037/0003-066X.54.7.493
77. Peter M. Gollwitzer and Paschal Sheeran. 2006. Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. In *Advances in Experimental Social Psychology*. Academic Press, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1
78. Google. YouTube Leanback offers effortless viewing. Retrieved September 12, 2020 from https://youtube.googleblog.com/2010/07/youtube-leanback-offers-effortless.html
79. Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–14. https://doi.org/10.1145/3173574.3174108
80. Greiffenhagen and Reeves. 2013. Is Replication Important for HCI? *RepliCHI*.
81. Marco Gui, Luca Stanca, and Others. 2009. *Television viewing, satisfaction and happiness: Facts and fiction*. repec.dems.unimib.it. Retrieved from http://repec.dems.unimib.it/repec/pdf/mibwpaper167.pdf
82. Keach Hagey and Jeff Horwitz. 2021. Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead. *WSJ Online*. Retrieved June 21, 2022 from https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215
83. Jonathan Haidt. 2006. *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom*. Basic Books. Retrieved from https://play.google.com/store/books/details?id=gHEv9yzj_a4C
84. Kevin A. Hallgren. 2012. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in quantitative methods for psychology* 8, 1: 23–34. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402032/
85. Zac Hall. 2019. Apple makes privacy extremely relatable in fun new iPhone ad - 9to5Mac. Retrieved September 13, 2020 from https://9to5mac.com/2019/03/14/iphone-privacy-ad/
86. Jaron Harambam, Dimitrios Bountouridis, Mykola Makhortykh, and Joris van Hoboken. 2019. Designing for the Better by Taking Users into Account: A Qualitative Evaluation of User Control Mechanisms in (News) Recommender Systems. In *Proceedings of the 13th ACM Conference on Recommender Systems* (RecSys '19), 69–77. https://doi.org/10.1145/3298689.3347014
87. Joshua Hardwick. 2018. Dwell Time: Is It Really a Ranking Factor? (And If So, Should You Care?). *SEO Blog by Ahrefs*. Retrieved March 28, 2019 from https://ahrefs.com/blog/dwell-time/
88. J. Hari. 2022. Stolen Focus: Why You Can't Pay Attention.
89. Ellie Harmon and Melissa Mazmanian. 2013. Stories of the Smartphone in Everyday Discourse: Conflict, Tension & Instability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 1051–1060. https://doi.org/10.1145/2470654.2466134
90. Steven C. Hayes. 2004. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior therapy* 35, 4: 639–665. Retrieved from http://www.sciencedirect.com/science/article/pii/S0005789404800133
91. Jennifer Healey. 2011. Recording Affect in the Field: Towards Methods and Metrics for Improving Ground Truth Labels. In *Affective Computing and Intelligent Interaction*, 107–116. https://doi.org/10.1007/978-3-642-24600-5_14
92. Hal E. Hershfield, Daniel G. Goldstein, William F. Sharpe, Jesse Fox, Leo Yeykelis, Laura L. Carstensen, and Jeremy N. Bailenson. 2011. INCREASING SAVING BEHAVIOR THROUGH AGE-PROGRESSED RENDERINGS OF THE FUTURE SELF. *JMR, Journal of marketing research* 48: S23–S37. https://doi.org/10.1509/jmkr.48.SPL.S23
93. Alexis Hiniker, Sharon S. Heung, Sungsoo (ray) Hong, and Julie A. Kientz. 2018. Coco's Videos: An Empirical Investigation of Video-Player Design Features and Children's Media Use. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–13.

https://doi.org/10.1145/3173574.3173828
94. 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*, 4746–4757. https://doi.org/10.1145/2858036.2858403
95. 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* (IDC '17), 85–95. https://doi.org/10.1145/3078072.3079752
96. 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*, 634–645. Retrieved from http://dl.acm.org/citation.cfm?id=2971762
97. A. Hiniker, H. Suh, S. Cao, and J. A. Kientz. 2016. Screen time tantrums: how families manage screen media experiences for toddlers and preschoolers. *of the 2016 CHI conference on* …. Retrieved from https://dl.acm.org/doi/abs/10.1145/2858036.2858278?casa_token=VJoUGY85dhIAAAAA:1IkfQwRP7VkZ9SnJHTqGnbiqbLPvohAbWaeAVxpHr1qsuksfx4Pnb-kLnYTDvfL6HZ1P8f1o3cEhFQ
98. 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: 1318–1335. https://doi.org/10.1037/a0026545
99. Kai T. Horstmann and Matthias Ziegler. 2020. Assessing personality states: What to consider when constructing personality state measures. *European journal of personality* 34, 6: 1037–1059. https://doi.org/10.1002/per.2266
100. Rosalind Hursthouse. 2003. Virtue Ethics. Retrieved April 5, 2022 from https://meinong.stanford.edu/entries/ethics-virtue/
101. Michael Inzlicht and Malte Friese. 2020. Willpower Is Overrated. https://doi.org/10.31234/osf.io/rfk6z
102. Riitta Jääskeläinen. 2010. Think-aloud protocol. *Handbook of translation studies* 1: 371–374. Retrieved from https://books.google.com/books?hl=en&lr=&id=sBVGAYCh_9AC&oi=fnd&pg=PA371&dq=Riitta+J%C3%A4%C3%A4skel%C3%A4inen+2010+Think-aloud+protocol.&ots=Qn8NYacfXD&sig=9gAnKpZ6vdVbGHapKTfo74b0MtE
103. Jeff Johnson. 2020. *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines*. Morgan Kaufmann. Retrieved from https://play.google.com/store/books/details?id=_dLVDwAAQBAJ
104. D. Kahneman. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux. Retrieved from https://books.google.com/books?id=SHvzzuCnuv8C
105. Anya Kamenetz. 2018. *The art of screen time: How your family can balance digital media and real life*. Hachette UK.
106. Barbara K. Kaye. 1998. Uses and gratifications of the World Wide Web: From couch potato to web potato. *New Jersey Journal of Communication* 6, 1: 21–40. https://doi.org/10.1080/15456879809367333
107. Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 4521–4532. https://doi.org/10.1145/2858036.2858465
108. 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: 29. https://doi.org/10.1145/3314403
109. J. Kim, J. Park, H. Lee, M. Ko, and U. Lee. 2019. LocknType: Lockout task intervention for discouraging smartphone app use. *of the 2019 CHI conference on …*. Retrieved from https://dl.acm.org/doi/abs/10.1145/3290605.3300927
110. 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* (CHI '16), 272–283. https://doi.org/10.1145/2858036.2858428
111. 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
112. Minsam Ko, Seungwoo Choi, Subin Yang, Joonwon Lee, and Uichin Lee. 2015. FamiLync: Facilitating Participatory Parental Mediation of Adolescents' Smartphone Use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15), 867–878. https://doi.org/10.1145/2750858.2804283
113. 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*, 998–1010. Retrieved from https://dl.acm.org/citation.cfm?id=2858568
114. 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* (CSCW '15), 1235–1245. https://doi.org/10.1145/2675133.2675244
115. Ksenia Koroleva, Hanna Krasnova, Natasha F. Veltri, and Oliver Günther. 2011. It's all about networking! Empirical investigation of social capital formation on social network sites. In *Proceedings of the Thirty Second International Conference on Information Systems (ICIS)*, 20. https://doi.org/10.7892/boris.47120
116. Geza Kovacs, Drew Mylander Gregory, Zilin Ma, Zhengxuan Wu, Golrokh Emami, Jacob Ray, and Michael S. Bernstein. 2019. Conservation of Procrastination: Do Productivity Interventions Save Time Or Just Redistribute It? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 330:1–330:12. https://doi.org/10.1145/3290605.3300560
117. 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
118. S. S. Krishnan and R. K. Sitaraman. 2013. Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs. *IEEE/ACM Transactions on Networking* 21, 6: 2001–2014. https://doi.org/10.1109/TNET.2013.2281542
119. Ethan Kross, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablack, John Jonides, and Oscar Ybarra. 2013. Facebook use predicts declines in subjective well-being in young adults. *PloS one* 8, 8: e69841. https://doi.org/10.1371/journal.pone.0069841
120. A. Kuznetsova, P. B. Brockhoff, Rhb Christensen - rproject. org/web/packages/lmerTest …, and 2015. 2015. lmerTest: tests in linear mixed effects models. R package version 2.0-20.

121. Klodiana Lanaj, Russell E. Johnson, and Christopher M. Barnes. 2014. Beginning the workday yet already depleted? Consequences of late-night smartphone use and sleep. *Organizational Behavior and Human Decision Processes 124*, 11–23. https://doi.org/10.1016/j.obhdp.2014.01.001
122. J. R. Landis and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1: 159–174. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/843571
123. 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: 101:1–101:22. https://doi.org/10.1145/3274370
124. Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2011. Mining Behavioral Economics to Design Persuasive Technology for Healthy Choices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), 325–334. https://doi.org/10.1145/1978942.1978989
125. Louis Leung and Ran Wei. 2000. More Than Just Talk on the Move: Uses and Gratifications of the Cellular Phone. *Journalism & mass communication quarterly* 77, 2: 308–320. https://doi.org/10.1177/107769900007700206
126. David M. Levy. 2016. *Mindful Tech: How to Bring Balance to Our Digital Lives*. Yale University Press. Retrieved from https://market.android.com/details?id=book-jOs_CwAAQBAJ
127. Richard J. Light. 1971. Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological bulletin* 76, 5: 365. https://doi.org/10.1037/h0031643
128. Hannah Limerick, David Coyle, and James W. Moore. 2014. The experience of agency in human-computer interactions: a review. *Frontiers in human neuroscience* 8: 643. https://doi.org/10.3389/fnhum.2014.00643
129. Hannah Limerick, James W. Moore, and David Coyle. 2015. Empirical Evidence for a Diminished Sense of Agency in Speech Interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 3967–3970. https://doi.org/10.1145/2702123.2702379
130. Markus Löchtefeld, Matthias Böhmer, and Lyubomir Ganev. 2013. AppDetox: Helping Users with Mobile App Addiction. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (MUM '13), 43:1–43:2. https://doi.org/10.1145/2541831.2541870
131. Danielle Lottridge, Eli Marschner, Ellen Wang, Maria Romanovsky, and Clifford Nass. 2012. Browser Design Impacts Multitasking. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting* 56, 1: 1957–1961. https://doi.org/10.1177/1071181312561289
132. Kai Lukoff. 2021. SwitchTube: How Control Over the Interface of a Video App Affects Satisfaction and Personal Goals [OSF experiment preregistration]. https://doi.org/10.17605/OSF.IO/SEVFD
133. Kai Lukoff, Alexis Hiniker, Colin M. Gray, Arunesh Mathur, and Shruthi Chivukula. 2021. What Can CHI Do About Dark Patterns? In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411763.3441360
134. 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)*, 5.
135. Kai Lukoff, Ulrik Lyngs, Stefania Gueorguieva, Erika S. Dillman, Alexis Hiniker, and Sean A. Munson. 2020. From Ancient Contemplative Practice to the App Store: Designing a Digital Container for Mindfulness. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 1551–1564.

https://doi.org/10.1145/3357236.3395444
136. 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
137. Kai Lukoff, Carol Moser, and Sarita Schoenebeck. 2017. Gender Norms and Attitudes about Childcare Activities Presented on Father Blogs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 4966–4971. https://doi.org/10.1145/3025453.3025767
138. Kai Lukoff and Sean Munson. 2019. Digital wellbeing is way more than just reducing screen time. *UX Collective*. Retrieved October 1, 2020 from https://uxdesign.cc/digital-wellbeing-more-than-just-reducing-screen-time-46223db9f057
139. Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1: 1–26. https://doi.org/10.1145/3191754
140. U. Lyngs, R. Binns, and M. Van Kleek. 2018. So, Tell Me What Users Want, What They Really, Really Want! *Extended Abstracts of the*. Retrieved from https://dl.acm.org/citation.cfm?id=3188397
141. Ulrik Lyngs. 2018. A Cognitive Design Space for Supporting Self-Regulation of ICT Use. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA '18), 1–6. https://doi.org/10.1145/3170427.3180296
142. 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*: 102869. https://doi.org/10.1016/j.ijhcs.2022.102869
143. 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*. https://doi.org/10.1145/3290605.3300361
144. Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Leek, and Nigel Shadbolt. 2019. Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. *CHI 2019*. https://doi.org/10.1145/3290605.3300361
145. 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
146. Kate Magsamen-Conrad, John Dowd, Mohammad Abuljadail, Saud Alsulaiman, and Adnan Shareefi. 2015. Life-Span Differences in the Uses and Gratifications of Tablets: Implications for Older Adults. *Computers in human behavior* 52: 96–106. https://doi.org/10.1016/j.chb.2015.05.024
147. Claudia Marino, Gianluca Gini, Alessio Vieno, and Marcantonio M. Spada. 2018. A comprehensive meta-analysis on Problematic Facebook Use. *Computers in human behavior* 83: 262–277. https://doi.org/10.1016/j.chb.2018.02.009
148. Gloria Mark, Mary Czerwinski, and Shamsi T. Iqbal. 2018. Effects of Individual Differences in Blocking Workplace Distractions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 92:1–92:12. https://doi.org/10.1145/3173574.3173666
149. Lucas Matney. 2017. YouTube has 1.5 billion logged-in monthly users watching a ton of mobile video. *TechCrunch*. Retrieved from

http://techcrunch.com/2017/06/22/youtube-has-1-5-billion-logged-in-monthly-users-watching-a-ton-of-mobile-video/
150. Melissa Mazmanian, Joanne Yates, and Wanda Orlikowski. 2006. UBIQUITOUS EMAIL: INDIVIDUAL EXPERIENCES AND ORGANIZATIONAL CONSEQUENCES OF BLACKBERRY USE. *Proceedings: a conference of the American Medical Informatics Association / ... AMIA Annual Fall Symposium. AMIA Fall Symposium* 2006, 1: D1–D6. https://doi.org/10.5465/ambpp.2006.27169074
151. John McAlaney, Manal Aldhayan, Mohamed Basel Almourad, Sainabou Cham, and Raian Ali. 2020. On the Need for Cultural Sensitivity in Digital Wellbeing Tools and Messages: A UK-China Comparison. In *Trends and Innovations in Information Systems and Technologies*, 723–733. https://doi.org/10.1007/978-3-030-45691-7_68
152. John McAlaney, Manal Aldhayan, Mohamed Basel Almourad, Sainabou Cham, and Raian Ali. 2020. Predictors of Acceptance and Rejection of Online Peer Support Groups as a Digital Wellbeing Tool. In *Trends and Innovations in Information Systems and Technologies*, 95–107. https://doi.org/10.1007/978-3-030-45697-9_10
153. Tom McKay. 2019. Senators Introduce Bill to Stop "Dark Patterns" Huge Platforms Use to Trick Users. *Gizmodo*. Retrieved August 27, 2020 from https://gizmodo.com/senators-introduce-bill-to-stop-dark-patterns-huge-plat-1833929276
154. Sean M. McNee, Shyong K. Lam, Catherine Guetzlaff, Joseph A. Konstan, and John Riedl. 2003. Confidence displays and training in recommender systems. In *Proc. INTERACT*, 176–183. Retrieved from https://books.google.com/books?hl=en&lr=&id=PTg0fVYqgCcC&oi=fnd&pg=PA176&dq=low+confidence+%22recommender+system%22&ots=ObJIxzmCAZ&sig=0uf4-fGfAJyBGbfmCDJE0zpPsGU
155. Sean M. McNee, John Riedl, and Joseph A. Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '06), 1097–1101. https://doi.org/10.1145/1125451.1125659
156. Elisa D. Mekler and Kasper Hornbæk. 2016. Momentary Pleasure or Lasting Meaning?: Distinguishing Eudaimonic and Hedonic User Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 4509–4520. https://doi.org/10.1145/2858036.2858225
157. Elisa D. Mekler and Kasper Hornbæk. A Framework for the Experience of Meaning in Human-Computer Interaction. In *CHI 19*. https://doi.org/10.1145/3290605.3300455
158. Janet Metcalfe and Matthew Jason Greene. 2007. Metacognition of agency. *Journal of experimental psychology. General* 136, 2: 184–199. https://doi.org/10.1037/0096-3445.136.2.184
159. Sandra Metts, Susan Sprecher, William R. Cupach, Barbara M. Montgomery, and Steve Duck. 1991. Retrospective self-reports. *Studying interpersonal interaction*: 162–178. Retrieved from https://books.google.com/books?hl=en&lr=&id=RxdkR7fC5gEC&oi=fnd&pg=PA162&dq=Retrospective+Self-Reports&ots=l4qSyRf2P1&sig=oGLlldHa_VrW9oDwtKr81NVl0n8
160. Brian T. Miller and Mark D'Esposito. 2005. Searching for "the Top" in Top-Down Control. *Neuron* 48, 4: 535–538. https://doi.org/10.1016/j.neuron.2005.11.002
161. E. K. Miller and J. D. Cohen. 2001. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* 24: 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167
162. David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151, 4: 264–9, W64. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

163. Alberto Monge Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 386:1–386:14. https://doi.org/10.1145/3290605.3300616
164. James W. Moore. 2016. What Is the Sense of Agency and Why Does it Matter? *Frontiers in psychology* 7: 1272. https://doi.org/10.3389/fpsyg.2016.01272
165. Carol Moser, Sarita Y. Schoenebeck, and Paul Resnick. 2019. Impulse Buying: Design Practices and Consumer Needs. https://doi.org/10.1145/3290605.3300472
166. Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. A manifesto for reproducible science. *Nature human behaviour* 1: 0021. https://doi.org/10.1038/s41562-016-0021
167. Inez Myin-Germeys and Peter Kuppens. 2021. *The Open Handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies*. Center for Research on Experience Sampling and Ambulatory Methods Leuven (REAL).
168. Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. 2020. Dark Patterns: Past, Present, and Future: The Evolution of Tricky User Interfaces. *Queueing Systems. Theory and Applications* 18, 2: 67–92. https://doi.org/10.1145/3400899.3400901
169. John von Neumann and Oskar Morgenstern. 2007. *Theory of Games and Economic Behavior (Commemorative Edition)*. Princeton University Press. Retrieved from https://market.android.com/details?id=book-_aIGYI-jGEcC
170. C. Newport. 2016. Deep work: Rules for focused success in a distracted world.
171. Jakob Nielsen. 1994. 10 Heuristics for User Interface Design: Article by Jakob Nielsen. *Nielsen Norman Group*. Retrieved February 7, 2020 from https://www.nngroup.com/articles/ten-usability-heuristics/
172. D. A. Norman, Timothy Shallice, and Others. 2000. Attention to action: Willed and automatic control of behaviour. Retrieved from https://iris.sissa.it/handle/20.500.11767/30584
173. NPR. 2015. Episode 653: The Anti-Store. *NPR*. Retrieved from https://www.npr.org/sections/money/2015/09/25/443519599/episode-653-the-anti-store
174. Ted O'Donoghue and Matthew Rabin. 1999. Doing It Now or Later. *The American economic review* 89, 1: 103–124. https://doi.org/10.1257/aer.89.1.103
175. Ted O'Donoghue and Matthew Rabin. 2015. Present Bias: Lessons Learned and to Be Learned. *The American economic review* 105, 5: 273–279. https://doi.org/10.1257/aer.p20151085
176. Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. 2018. Good vibrations: can a digital nudge reduce digital overload? In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 4. https://doi.org/10.1145/3229434.3229463
177. Open Science Collaboration. 2015. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science* 349, 6251: aac4716. https://doi.org/10.1126/science.aac4716
178. Will Oremus. 2016. Who controls your Facebook feed. *Slate 27*.
179. Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing* 16, 1: 105–114. https://doi.org/10.1007/s00779-011-0412-2
180. Erica Pandey. 2017. Sean Parker: Facebook was designed to exploit human ``vulnerability''. Retrieved September 15, 2020 from https://www.axios.com/sean-parker-facebook-exploits-a-vulnerability-in-humans-2507917325.html
181. Joonyoung Park, Jin Yong Sim, Jaejeung Kim, Mun Yong Yi, and Uichin Lee. 2018. Interaction Restraint: Enforcing Adaptive Cognitive Tasks to Restrain Problematic User Interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA

’18), LBW559:1–LBW559:6. https://doi.org/10.1145/3170427.3188613
182. Joonyoung Park, Jin Yong Sim, Jaejeung Kim, Mun Yong Yi, and Uichin Lee. 2018. Interaction Restraint: Enforcing Adaptive Cognitive Tasks to Restrain Problematic User Interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA ’18), LBW559:1–LBW559:6. https://doi.org/10.1145/3170427.3188613
183. Nick Paumgarten. 2014. Up And Then Down. *The New Yorker*. Retrieved September 13, 2020 from https://www.newyorker.com/magazine/2008/04/21/up-and-then-down
184. Andrew Perrin and Monica Anderson. 2019. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. *Pew Research Center*. Retrieved September 14, 2020 from https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/
185. Pew Research Center. 2018. *How Teens and Parents Navigate Screen Time and Device Distractions*.
186. Laura Pina, Kael Rowan, Asta Roseway, Paul Johns, Gillian R. Hayes, and Mary Czerwinski. 2014. In situ cues for ADHD parenting strategies using mobile technology. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* (PervasiveHealth ’14), 17–24. https://doi.org/10.4108/icst.pervasivehealth.2014.254958
187. C. Pinder, J. Vermeulen, and B. R. Cowan. 2018. Digital Behaviour Change Interventions to Break and Form Habits. *ACM Transactions on*. Retrieved from https://dl.acm.org/citation.cfm?id=3196830
188. 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
189. John Raacke and Jennifer Bonds-Raacke. 2008. MySpace and Facebook: applying the uses and gratifications theory to exploring friend-networking sites. *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society* 11, 2: 169–174. https://doi.org/10.1089/cpb.2007.0056
190. Ahmad Rahmati, Chad Tossell, Clayton Shepard, Philip Kortum, and Lin Zhong. 2012. Exploring iPhone Usage: The Influence of Socioeconomic Differences on Smartphone Adoption, Usage and Usability. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services* (MobileHCI ’12), 11–20. https://doi.org/10.1145/2371574.2371577
191. Daniel Read, George Loewenstein, and Shobana Kalyanaraman. 1999. Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic. *Journal of behavioral decision making* 12, 4: 257–273. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1099-0771(199912)12:4%3C257::AID-BDM327%3E3.0.CO;2-6
192. D. Read and van Leeuwen B. 1998. Predicting Hunger: The Effects of Appetite and Delay on Choice. *Organizational behavior and human decision processes* 76, 2: 189–205. https://doi.org/10.1006/obhd.1998.2803
193. Paul Resnick and Hal R. Varian. 1997. Recommender systems. *Communications of the ACM* 40, 3: 56–58. Retrieved from http://dl.acm.org/citation.cfm?id=245121
194. Alberto Monge Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. https://doi.org/10.1145/3290605.3300616
195. Kevin Roose. 2019. The making of a YouTube radical. *The New York times*. Retrieved from https://static01.nyt.com/images/2019/06/09/nytfrontpage/scan.pdf
196. Larry D. Rosen, L. Mark Carrier, and Nancy A. Cheever. 2013. Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in human behavior* 29, 3:

948–958. https://doi.org/10.1016/j.chb.2012.12.001
197. W. Ruan, Q. Z. Sheng, L. Yao, N. K. Tran, and Y. C. Yang. 2016. PreventDark: Automatically detecting and preventing problematic use of smartphones in darkness. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, 1–3. https://doi.org/10.1109/PERCOMW.2016.7457071
198. Alan M. Rubin. 1983. Television uses and gratifications: The interactions of viewing patterns and motivations. *Journal of broadcasting & electronic media* 27, 1: 37–51. https://doi.org/10.1080/08838158309386471
199. Alan M. Rubin. 1984. Ritualized and Instrumental Television Viewing. *The Journal of communication* 34, 3: 67–77. https://doi.org/10.1111/j.1460-2466.1984.tb02174.x
200. Alan M. Rubin. 2009. Uses and gratifications. *The SAGE handbook of media processes and effects*: 147–159. Retrieved from https://books.google.com/books?hl=en&lr=&id=CMO1aEWrzacC&oi=fnd&pg=PT155&dq=Uses+and+gratifications+rubin+2009+sage+handbook&ots=sBmAUl_iwG&sig=QAxw2CcA3oY9yVZd-2BROIGebnw
201. Alan M. Rubin and Elizabeth M. Perse. 1987. Audience activity and soap opera involvement A uses and effects investigation. *Human communication research* 14, 2: 246–268. https://doi.org/10.1111/j.1468-2958.1987.tb00129.x
202. Thomas E. Ruggiero. 2000. Uses and Gratifications Theory in the 21st Century. *Mass Communication and Society* 3, 1: 3–37. https://doi.org/10.1207/S15327825MCS0301_02
203. James A. Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6: 1161. https://doi.org/10.1037/h0077714
204. Richard M. Ryan and Edward L. Deci. 2006. Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of personality* 74, 6: 1557–1586. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6494.2006.00420.x?casa_token=MZuzC4Br_U4AAAAA:ErU7WbByAWUFcoh2N_5TIRqe7jhVXe6V8Z0-pWB8gbb-ZZ3I8xz_qrdAtePASmiTFBWb2COF6sX4BQ
205. Christina Sagioglou and Tobias Greitemeyer. 2014. Facebook's emotional consequences: Why Facebook causes a decrease in mood and why people still use it. *Computers in human behavior* 35: 359–363. Retrieved from http://www.sciencedirect.com/science/article/pii/S0747563214001241
206. Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-scale assessment of mobile notifications. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 3055–3064. https://doi.org/10.1145/2556288.2557189
207. Saleema Amershi Microsoft Seattle, Dan Weld University of Washington & Microsoft Research, Seattle, WA, USA, Mihaela Vorvoreanu Microsoft Redmond, Adam Fourney Microsoft Redmond, Besmira Nushi Microsoft Redmond, Penny Collisson Microsoft Redmond, Jina Suh Microsoft Remond, Shamsi Iqbal Microsoft Redmond, Paul N. Bennett Microsoft, Redmond, WA, USA, Kori Inkpen Microsoft Redmond, Jaime Teevan Microsoft Redmond, Ruth Kikin-Gil Microsoft Redmond, and Eric Horvitz Microsoft Redmond. Guidelines for Human-AI Interaction | Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Retrieved January 27, 2020 from https://dl.acm.org/doi/10.1145/3290605.3300233
208. Mohammad Salehan and Arash Negahban. 2013. Social networking on smartphones: When mobile phones become addictive. *Computers in human behavior* 29, 6: 2632–2639. https://doi.org/10.1016/j.chb.2013.07.003
209. Blair Saunders, Marina Milyavskaya, Alexander Etz, Daniel Randles, and Michael Inzlicht.

2018. Reported Self-control is not Meaningfully Associated with Inhibition-related Executive Function: A Bayesian Analysis. *Collabra. Psychology* 4, 1: 39. https://doi.org/10.1525/collabra.134
210. Thomas C. Schelling. 1984. Self-Command in Practice, in Policy, and in a Theory of Rational Choice. *The American economic review* 74, 2: 1–11. Retrieved from http://www.jstor.org/stable/1816322
211. Markus Schlosser. 2019. Agency. *The Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/archives/win2019/entries/agency/
212. Sarita Yardi Schoenebeck. 2014. Giving up Twitter for Lent: how and why we take breaks from social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 773–782. https://doi.org/10.1145/2556288.2556983
213. Michael Schrage. 1996. Cultures of prototyping. *Bringing design to software* 4, 1: 1–11. Retrieved from https://pdfs.semanticscholar.org/c399/6e2738e52ea22c83ef662ab118f68b82eba2.pdf
214. Natasha Dow Schüll. 2012. *Addiction by Design*. Princeton University Press. https://doi.org/10.1515/9781400834655
215. N. D. Schüll. 2012. *Addiction by Design: Machine Gambling in Las Vegas*. Princeton University Press. Retrieved from https://books.google.com/books?id=_Vsk6EXc1_4C
216. Ralf Schwarzer. 2008. Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. *Applied psychology = Psychologie appliquee* 57, 1: 1–29. https://doi.org/10.1111/j.1464-0597.2007.00325.x
217. John Scott. 2000. Rational choice theory. *Understanding contemporary society: Theories of the present* 129: 671–685. Retrieved from https://books.google.com/books?hl=en&lr=&id=QaUgne7fgYUC&oi=fnd&pg=PA126&dq=rational+choice+theory&ots=2ANYUrj14k&sig=KvO955C1lKgF_1Td9ukgJM3vjOQ
218. Nick Seaver. 2018. Captivating algorithms: Recommender systems as traps. *Journal of Material Culture*: 1359183518820366. https://doi.org/10.1177/1359183518820366
219. 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: 1–19. https://doi.org/10.1016/j.entcom.2015.07.002
220. W. J. Severin and J. W. Tankard. 1997. Cognitive consistency and mass communication. *Communication Theories: Origins, Methods, and Uses in Mass Media*: 159–177.
221. Choonsung Shin and Anind K. Dey. 2013. Automatically Detecting Problematic Use of Smartphones. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '13), 335–344. https://doi.org/10.1145/2493432.2493443
222. Ben Shneiderman and Catherine Plaisant. 2004. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley.
223. Jonathan Sillito. Saturate App: Simple Collaborative Analysis. Retrieved from http://www.saturateapp.com/
224. L. Silver, A. Smith, C. Johnson, K. Taylor, J. Jiang, A. Monica, and L. Rainie. 2019. Use of smartphones and social media is common across most emerging economies. *Pew Research Center*. Retrieved from https://www.pewresearch.org/internet/2019/03/07/use-of-smartphones-and-social-media-is-common-across-most-emerging-economies/
225. Joan E. Solsman. 2018. Ever get caught in an unexpected hourlong YouTube binge? Thank YouTube AI for that. *CNET*. Retrieved May 1, 2020 from https://www.cnet.com/news/youtube-ces-2018-neal-mohan/
226. Todd Spangler. YouTube Tops 20 Million Paying Subscribers, YouTube TV Has Over 2 Million

Customers. Retrieved August 26, 2020 from https://variety.com/2020/digital/news/youtube-tops-20-million-paying-subscribers-youtube-tv-has-over-2-million-customers-1203491228/
227. Nick Statt. 2016. Flowstate is a writing app that will delete everything if you stop typing. Retrieved August 10, 2020 from https://www.theverge.com/2016/1/28/10853534/flowstate-writing-app-mac-ios-delete-everything
228. Universiti Putra Steven Eric Krauss. 2005. Research Paradigms and Meaning Making: A Primer. *The Qualitative Report* 10, 4: 758–770. Retrieved from http://nsuworks.nova.edu/tqr/vol10/iss4/7/
229. S. Shyam Sundar and Sampada S. Marathe. 2010. Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage. *Human communication research* 36, 3: 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x
230. Matthis Synofzik, Gottfried Vosgerau, and Albert Newen. 2008. Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and cognition* 17, 1: 219–239. https://doi.org/10.1016/j.concog.2007.03.010
231. Adam Tapal, Ela Oren, Reuven Dar, and Baruch Eitam. 2017. The Sense of Agency Scale: A Measure of Consciously Perceived Control over One's Mind, Body, and the Immediate Environment. *Frontiers in psychology* 8: 1552. https://doi.org/10.3389/fpsyg.2017.01552
232. Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Springer.
233. Jonathan A. Tran, Katherine S. Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. In *CHI '19*. https://doi.org/10.1145/3290605.3300542
234. 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* (CHI '19), 1–14. https://doi.org/10.1145/3290605.3300542
235. Amos Tversky and Daniel Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics* 106, 4: 1039–1061. Retrieved from https://academic.oup.com/qje/article-abstract/106/4/1039/1873382
236. United States Census Bureau. QuickFacts: United States. Retrieved August 14, 2020 from https://www.census.gov/quickfacts/fact/table/US/PST045219
237. Elizabeth Van Couvering. 2007. Is Relevance Relevant? Market, Science, and War: Discourses of Search Engine Quality. *Journal of computer-mediated communication: JCMC* 12, 3: 866–887. https://doi.org/10.1111/j.1083-6101.2007.00354.x
238. Philippe Verduyn, David Seungjae Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross. 2015. Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of experimental psychology. General* 144, 2: 480–488. https://doi.org/10.1037/xge0000057
239. Robert A. White. 1994. Audience "Interpretation" of Media: Emerging perspectives. *Communication research trends* 14, 3: 1–39. Retrieved May 16, 2022 from https://dialnet.unirioja.es/servlet/articulo?codigo=7803996
240. James Williams. 2018. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge University Press.
241. Max L. Wilson, Ed H. Chi, Stuart Reeves, and David Coyle. 2014. RepliCHI: the workshop II. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14), 33–36. https://doi.org/10.1145/2559206.2559233
242. Max L. L. Wilson, Paul Resnick, David Coyle, and Ed H. Chi. 2013. RepliCHI: the workshop. In

CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13), 3159–3162. https://doi.org/10.1145/2468356.2479636
243. 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: 43. https://doi.org/10.1186/1479-5868-8-43
244. Tim Wu. 2016. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Vintage. Retrieved from https://www.amazon.com/Attention-Merchants-Scramble-Inside-Heads-ebook/dp/B01AEPSWB4/ref=sr_1_1?crid=XBK4L7OF1VP0&keywords=the+attention+merchants+tim+wu&qid=1553794816&s=gateway&sprefix=the+attention+merchant%2Caps%2C191&sr=8-1
245. Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, and Shobha Venkataraman. 2011. Identifying Diverse Usage Behaviors of Smartphone Apps. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference* (IMC '11), 329–344. https://doi.org/10.1145/2068816.2068847
246. 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), 1–17. https://doi.org/10.1145/3491102.3517476
247. José P. Zagal, Staffan Björk, and Chris Lewis. 2013. Dark Patterns in the Design of Games. In *Foundations of Digital Games 2013*. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2:1043332
248. 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* (CHI '22), 1–19. https://doi.org/10.1145/3491102.3517722
249. S. Zika, K. Chamberlain - British journal of psychology, and 1992. 1992. On the relation between meaning in life and psychological well-being. *Wiley Online Library*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.2044-8295.1992.tb02429.x/full
250. John Zimmerman and Jodi Forlizzi. 2014. Research Through Design in HCI. In *Ways of Knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (eds.). Springer New York, New York, NY, 167–189. https://doi.org/10.1007/978-1-4939-0378-8_8
251. 2014. Value is created by doing. *Sam Altman*. Retrieved March 4, 2018 from https://blog.samaltman.com/value-is-created-by-doing
252. 2014. Some Dark Patterns now illegal. *90 Percent Of Everything*. Retrieved April 8, 2019 from https://www.90percentofeverything.com/2014/08/26/some-dark-patterns-now-illegal-in-uk-interview-with-heather-burns/
253. 2021. Mobile Fact Sheet. *Pew Research Center: Internet, Science & Tech*. Retrieved April 4, 2022 from https://www.pewresearch.org/internet/fact-sheet/mobile/
254. Deceptive Design – formerly darkpatterns.org. Retrieved April 5, 2022 from http://darkpatterns.org
255. YouTube Regrets — Mozilla Foundation. *Mozilla Foundation*. Retrieved May 1, 2020 from https://foundation.mozilla.org/en/campaigns/youtube-regrets/
256. google-play-scraper. *PyPI*. Retrieved June 21, 2022 from https://pypi.org/project/google-play-scraper/
257. Digital Services Act: Council and European Parliament provisional agreement for making the internet a safer space for European citizens. Retrieved May 15, 2022 from https://www.consilium.europa.eu/en/press/press-releases/2022/04/23/digital-services-act-c

ouncil-and-european-parliament-reach-deal-on-a-safer-online-space/