

# Monitoring Screen Time or Redesigning It?

Two Approaches to Supporting Intentional Social Media Use

Mingrui “Ray” Zhang  
The Information School  
University of Washington  
Seattle, USA  
mingrui@uw.edu

Kai Lukoff  
Human Centered Design &  
Engineering  
University of Washington  
Seattle, USA  
kai1@uw.edu

Raveena Rao  
The Information School  
University of Washington  
Seattle, USA  
rraveena@uw.edu

Amanda Baughan  
Paul G. Allen School of Computer  
Science & Engineering  
University of Washington  
Seattle, USA  
baughan@cs.washington.edu

Alexis Hiniker  
The Information School  
University of Washington  
Seattle, USA  
alexisr@uw.edu

## ABSTRACT

Existing designs helping people manage their social media use include: 1) external supports that monitor and limit use; 2) internal supports that change the interface itself. Here, we design and deploy Chirp, a mobile Twitter client, to independently examine how users experience external and internal supports. To develop Chirp, we identified 16 features that influence users’ sense of agency on Twitter through a survey of 129 participants and a design workshop. We then conducted a four-week within-subjects deployment with 31 participants. Our internal supports (including features to filter tweets and inform users when they have exhausted new content) significantly increased users’ sense of agency, while our external supports (a usage dashboard and nudges to close the app) did not. Participants valued our internal supports and said that our external supports were for “other people.” Our findings suggest that design patterns promoting agency may serve users better than screen time tools.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**;  
**Human computer interaction (HCI)**.

## KEYWORDS

sense of agency, mobile phone, digital wellbeing, social media, twitter

### ACM Reference Format:

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)*, April 29-May 5, 2022, New Orleans, LA,



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI ’22, April 29-May 5, 2022, New Orleans, LA, USA  
© 2022 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-9157-3/22/04.  
<https://doi.org/10.1145/3491102.3517722>

USA. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3491102.3517722>

## 1 INTRODUCTION

The attention economy [47] depends on serving advertisements to users, incentivizing designers to create interfaces that draw users back as often possible and keep them captivated as long as possible. The resulting design strategies are often successful in capturing users’ attention but are not always aligned with human well-being. People express widespread frustration with designs like click-bait titles, infinite scrolling, and intrusive notifications, which leave them feeling like the interface is working to undermine their decision-making and manipulate how they spend their time [59]. Many people say that they routinely find themselves spending more time than they would like on specific digital experiences [33] and that they struggle to change their habits [9].

A key reason for users’ frustration is the importance of *sense of agency* on people’s well-being. Maintaining a “*feeling of control over actions and their consequences*” [49] is a fundamental human need [51] and essential to living a fulfilling life [40]. An increase in a person’s sense of agency predicts increases in physical health [55], supportive relationships [16], and subjective well-being [36]. People consistently describe resistance to and frustration with interfaces they perceive to erode their agency [2, 3, 56, 59].

A large body of prior work in HCI has examined users’ sense of agency with respect to the apps and devices they use. These studies document that users routinely delete apps and accounts [54, 59], take extended breaks [56], or set up physical barriers [19] to reduce their use of digital experiences that they perceive to be wasting their time or manipulating them into endless scrolling. To support users in regaining their agency, researchers have designed and evaluated a number of systems to help users manage, monitor and set boundaries on their own usage. These include, for example, stand-alone timers [18], usage trackers [28], automated nudges to encourage the user to end an experience [52], tools to set goals to reduce usage [33], and lock-out mechanisms to make an experience inaccessible [38].

Although many of these interventions have been shown to reduce people’s use, effects tend to decline over time [34] and often disappear after the intervention is removed [52]. Furthermore, in practice, such tools are routinely abandoned [10]. Prior work also shows that people are reluctant to adopt or stick with such tools because they do not *want* to self-police their usage or lock themselves out of digital experiences that offer value [27, 34, 59]. This raises the question: what can designers do to support user agency, given that people are frustrated with their current behaviors but find outside supports unsatisfactory?

Prior work has proposed that to increase users’ sense of agency, designers could shift from designing *external* supports to designing *internal* supports [41]. That is, rather than reactively providing users with external tools to self-monitor and self-limit (i.e., external supports), designers might instead proactively make changes to the internal workings of their interface to avoid undermining users’ sense of agency in the first place (i.e., internal supports). External supports might include, for example, timers and locks, whereas internal supports might instead swap out an infinite feed for a curated set of items, or provide a toggle for turning off autoplay.

Currently, creating external supports is the dominant design approach to promoting user agency online [44, 48, 61]. But given the shortcomings users report in these screen time tools, we sought to examine the value and drawbacks of providing external versus internal supports. As a test case, we designed and evaluated external and internal supports specifically for the Twitter app, a social networking service for posting, reading, and interacting with short messages called “tweets.” Users of Twitter frequently report a loss of control [29, 56], and the current Twitter interfaces contain features that smartphone users report lead to compulsive use, such as infinite scrolling [59]. Specifically, to examine the idea that internal supports may be a useful alternative or complement to external ones, we ask:

**RQ1.** What existing features influence users’ sense of agency on Twitter?

**RQ2.** What internal and external supports might promote users’ sense of agency on Twitter?

**RQ3.** How, if at all, do these internal and external supports affect users’ sense of agency in practice?

**RQ4.** What differences, if any, do users experience between these external and internal supports?

To answer RQ1, we conducted a survey with 129 Twitter users asking what makes them feel most and least in control over how they use the Twitter mobile app. We then used the findings from this survey to seed a design workshop with four experts, brainstorming potential interface changes that could improve users’ sense of agency (RQ2) given the challenges participants expressed in the survey. Drawing on this formative work, we developed “Chirp,” an alternative Twitter mobile client.<sup>1</sup> We implemented six features that emerged during the workshop as potential avenues to promote user agency. This included both internal supports (a Feed Filter, Reading Progress Indicator, Custom Lists, and a Recommended Tweets Blocker) and external supports (a Usage Stats Page and a Time Limit Dialog). To evaluate users’ experiences with these designs (RQ3) and to examine the effectiveness of these internal

and external supports independently (RQ4), we conducted a four-week deployment study with 31 U.S. participants, collecting usage data through automated logs, experiential data through experience sampling method (ESM) prompts, and contextual, self-reflection data through weekly surveys and an exit interview.

We found that the internal supports we implemented increased users’ sense of agency, though external supports did not. Internal supports surgically targeted specific aspects of the user experience (such as re-reading old content or sifting through irrelevant tweets), and we found that increases in agency arose from changes to users’ behavior in the target scenarios. Although some users appreciated being able to view their usage data, most found our external supports only minimally helpful, and some people said these tools undermined their agency by adversarially inhibiting their usage instead of partnering with them to improve it. Many participants said that, in particular, the Time Limit Dialog, which prompted them to exit the app, was a nuisance. Interestingly, they often went on to say that it would likely be helpful for other people who struggle with self-control. These results suggest that the design community can improve users’ digital wellbeing by defining and evaluating internal supports and generalizing them into patterns that can be applied across social media platforms.

## 2 RELATED WORK

### 2.1 Frustration with Habitual Technology Use

Many people report frustration with their technology usage habits. They say that their phones distract them when they would prefer to focus [24], encourage compulsive habits [59], and lead them to behaviors they dislike but cannot change [33]. Some experiences produce more frustration than others, and many people report that social media and casual games are particularly likely to entice them into spending more time on task than they would like [18]. The construct of *lagging resistance* (in contrast to lagging adoption) refers to the common sentiment of wanting to cut back on the use of a particular app or system but not doing so just yet [1], reflecting people’s interest in—and struggle to—change their usage habits.

Some users successfully make changes to their usage habits but only by taking extreme steps to cut off their access to experiences they value. Prior work documents users avoiding social media sites they enjoy [2], locking themselves out of their accounts by handing their passwords over to a friend [1], deactivating their accounts [9], or deleting apps altogether [59]. This cold-turkey approach has costs for users and industry alike. The apps and devices people use often offer enormous benefits to their users, including increases in social capital [14], interpersonal connections [50], and subjective well-being [6]. Thus, it is well worth supporting users in findings ways to use systems they value at the cadence that is best for them. Scholars label the act of achieving this balance “digital wellbeing” [60].

Although users’ desire for digital wellbeing is well documented, this end goal remains elusive. In the present study, we seek to nudge this space forward by contributing to the body of empirical evidence on design decisions that are—and are not—likely to promote digital wellbeing.

<sup>1</sup>Source code is available on GitHub at <https://github.com/uclab/Chirp>

## 2.2 Frustration with Loss of Control on Twitter and Other Social Media Platforms

Twitter is one of many platforms that incites lagging resistance and user frustration. Twitter is a microblogging service where users, from ordinary people to celebrities, can broadcast short posts called tweets (currently limited to 280 characters each) and follow the broadcasts of individuals of their choosing. In 2006, Twitter began as a platform for sharing brief status updates with friends, but it has since expanded from this exclusively interpersonal focus to serve a wide variety of informational, cultural, and geopolitical needs [5]. Twitter is regularly used by companies to promote their brand [13], by activists to rally the public to a cause [20], by communities to share information during crises [57], by journalists to document breaking news, and by bad actors and special interests to spread misinformation and propaganda [15].

Like other social media platforms, Twitter leaves many end-users struggling to control their usage habits. For example, people say they wish they could give up Twitter for Lent (a 40-day religious period during which people choose to refrain from certain “vices”) but fail because resisting the urge to check Twitter is too daunting [56]. Twitter users also report that they find themselves saying “just a couple of minutes more” and lingering on the platform longer than they plan to [29]. Twitter shares many features—such as click-bait titles, infinite scrolling, and constant notifications—with other social platforms that users say undermine their intentions and lead to frustration. For example, Facebook users say they wish they could curb compulsive use [45], YouTube users describe recommendations leading them down rabbit holes [41], and Instagram users report trying and failing to cut down on their time in the app [30]. One goal of the current work is to examine which aspects of Twitter’s design users see as linked to compulsive habits and usage patterns they would like to change.

Of course, people also have intentions that drive them to use social media in the first place. For Twitter, the most common gratifications that users seek are obtaining information, connecting socially, and having fun [8, 11, 22]. On YouTube, users seek entertainment, information, and, to a lesser extent, social interaction and status [23]. This raises the question of how users might be empowered to still obtain their desired gratifications from these social media without sacrificing their sense of control and lapsing into compulsive use.

## 2.3 Designing for User Agency

A number of studies have examined how designers might support users in maintaining agency over their technology use and promoting their digital wellbeing [7]. Several investigations have demonstrated that tools to support users in curbing their use of technology can be effective. For example, Ko and colleagues successfully leveraged social support techniques to create an app for groups of college students to set limits on their collective phone use [33]. FamLync enabled parents and children to cooperate in setting limits on usage [31] and the AppDetox research project enables users to set rules to block specific apps [39]. Other research projects have sought to support users in cutting back in targeted ways. The PomodoLock research project supports users in resisting self-interruptions while trying to focus on a task by locking them

out of distraction experiences for short chunks of time [26]. Let’s FOCUS reduces technology-related distractions in classroom settings [25]. And Lock ‘n’ Lol supports users in self-regulating their use of technology while spending time with others in person [32].

Other work has brought together multiple solutions or comparative different approaches. MyTime is a composite intervention that leverages multiple forms of support simultaneously, including goal setting, passive information, and timers [18]. The HabitLab platform rotates through more than 20 different intervention approaches, demonstrating that this novelty improves effectiveness [34]. Still other work examines the tradeoffs between different interventions, showing, for example that reminders about goals are more annoying but less likely to make users anxious than interventions that suppress content [45].

## 2.4 Current Shortcomings of Designs for User Agency

Despite the many innovations to design for user agency and digital wellbeing, people’s struggles remain widespread. Prior work shows that users who adopt restrictive strategies for reducing their technology use gradually shift to move permissive ones [35] and other work shows that people often find such tools intrusive and annoying [45]. In a co-design study in which participants created tools to support their own usage goals, people first designed lockout mechanisms and other tools to help them self-police their use but then said they would be unlikely to adopt their own creations [59]. Other work finds that people appreciate digital wellbeing tools and believe they have promise, yet currently find them ineffective in practice [48].

This ambivalence suggests that people do not want to have adversarial relationships with the technologies they use. Today, the primary approach to designing for digital wellbeing is to arm users with tools to shut down their own engaged habits [44, 48, 61]. But users may well prefer not to be enticed into habits they dislike in the first place. Prior work found that removing autoplay and adding parental controls were equally effective in helping children to manage their YouTube use [17]. Consistent with this finding, prior work has posited that creating *internal supports* for user agency (that is, designs that make the original experience easier to manage) may be a useful complement to *external supports*, like the current ecosystem of screen time tools, that empower the user to lock themselves out of an experience they enjoy [41]. In this work, we take up the question of what internal and external supports have to offer by conducting an extensive design process to create a set of each for the Twitter platform and using an experimental field deployment to evaluate their effects.

## 3 DESIGN PROCESS

To examine how the current Twitter mobile app affects users’ sense of agency (RQ1), we first conducted a survey with 129 U.S. Twitter users. We then used the features that participants surfaced in the survey to create prompts for a design workshop. Using designs from the workshop, we identified a set of internal and external supports to evaluate in a novel Twitter client.

**Table 1: Demographics of survey participants**

Gender Identity	Man (55.0%), Woman (43.4%), Prefer not to say (1.6%)
Age	18-24 (7.8%), 25-34 (48.1%), 35-44 (24.8%), 45-54 (10.9%), 55+ (8.5%)
Education	High school (17.1%), Associate degree (12.4%), Bachelor’s degree (52.7%), Advanced degree (17.8%)
Household Income (USD)	<25K (10.1%), 25-50K (24.0%), 50-75K (24.0%), 75-125K (25.6%), >125K (14.7%), Prefer not to say (1.6%)
Race	White (76.0%), Asian (11.6%), Black (11.6%), Hispanic/Latino (5.4%), Other (0.8%)
Smartphone System	iOS (50.4%), Android (49.6%)
Average Daily Twitter Use	>2 Hours (9%), 1-2 Hours (23%), 10-60 Minutes (53%), 5-10 Minutes (16%)

**Table 2: The “most in control” section of the survey along with sample responses from a study participant. All participants also completed a second, complementary section of the survey where the text “most” was modified to “least.”**

	What are 3 things about the mobile app that lead you to feel most in control over how you spend your time on Twitter?	How does this thing make you feel most in control of how you spend your time on Twitter?
Thing 1	“Muting accounts”	“It allows me to be free to see what I want on my page without being disrupted by things I don’t wish to see or people I don’t like.” (P103)
Thing 2	“Direct messaging”	“Twitter gives you the option to only receive DMs from people you follow, and it helps to cut down on harassing or spam messages.” (P64)
Thing 3	“I can turn off notifications.”	“I can turn off notifications so I will not be distracted and log in when I had not planned to.” (P66)

### 3.1 Survey of Current Twitter Users

**Participants** We recruited 130 mobile Twitter users through Amazon’s Mechanical Turk platform, one of whom was excluded for incomplete responses. The remaining 129 participants were all located in the U.S. and had previously completed at least 1000 tasks on the Mechanical Turk platform with a task approval rating of at least 99%. Participants were required to own a smartphone with the Twitter client installed and to spend a minimum of 10 minutes per day using Twitter, with at least 10% of their total Twitter use occurring on their phone. Median Twitter use among participants was consistent with the average Twitter user, who used the platform for an average of 10 minutes per day in 2019 [53]. Participant demographics are shown in Table 1.

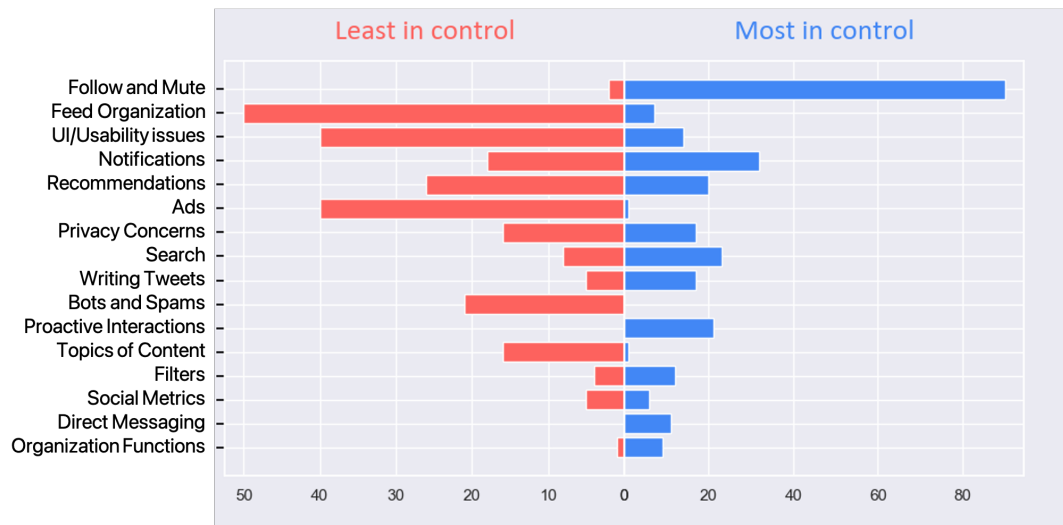
**Materials and Procedures** We created a survey modeled on a prior HCI survey study that examined users’ sense of agency when using YouTube [41]. The survey first screened for our inclusion criteria; eligible participants then continued to the remaining questions, which asked about demographics and participants’ perspectives on Twitter. To investigate how the design of Twitter influences users’ sense of agency (RQ1), the survey asked about the aspects of Twitter that make participants feel *most in control* and *least in control* when using the platform (Table 2). The order of these survey sections (most versus least) was randomized. We used the wording “feel in control” in keeping with prior work examining the sense of agency [46, 58]. Each participant listed six aspects of Twitter that influence their sense of agency when using the platform, three in response to the prompt asking what makes them feel most in control and three for least in control. Participants were compensated

\$4 for answering all questions (20 minutes). The complete survey can be found in Appendix A.

**Data Analysis** The first and second authors reviewed all 774 responses (six per participant) and conducted weekly meetings to develop initial codes, using thematic analysis [4]. We excluded 214 responses that did not include reflections on specific features. Of these, 89 were obviously not related to the question (including some responses that seemed to be generated by bots) (e.g., “I can walk for 60 minutes in the morning, and then build on that success daily”). An additional 41 lacked substantive content (e.g., “I choose when and where I log on to Twitter”), 33 did not mention a specific feature of the Twitter app (e.g., “Ease of use for fast response time”), and 51 contained only general feelings about the app (e.g., “I’m not truly worried about being addicted to online social media browsing but if doing it too much I can always delete the app”). To ensure that no response was mistakenly excluded, the first and third authors both independently coded the excluded responses with the four exclusion reasons. Any responses which were not excluded by both authors were discussed to reach an agreement, such that both authors agreed on all exclusions.

The coding process was conducted via a shared spreadsheet, and the initial codebook contained 33 Twitter features.<sup>2</sup> The first and second authors then applied the initial codes to 50 randomly selected responses (25 in-control responses and 25 out-of-control responses) and met to discuss. Interrater reliability was assessed using Cohen’s kappa, with  $\kappa = 0.80$ , indicating substantial agreement

<sup>2</sup>All first-round codes are presented in Appendix B



**Figure 1:** This diverging bar chart shows how many times the 16 features were mentioned as leading participants to feel *least* or *most* in control when using the Twitter mobile app. Features are sorted by the frequency with which they were mentioned by participants.

[37]. After the first round of coding, the entire research team met online and: 1) combined similar codes; 2) developed hierarchical subcodes for codes that appeared in many responses (for example, “notification mechanism” contained subcodes of “skimming notifications” and “notification settings”), and 3) merged similar codes into coherent themes. Our final codebook consisted of *eight themes* describing *24 different features (codes)* drawn from 560 responses (336 about features that make users feel most in control, 224 for least in control). The first two authors then conducted a second coding pass of all responses together using the revised codebook. The first author wrote a summary of the coding results with examples of each code and discussed the results with team members in weekly meetings. We focus on the 16 features that were mentioned more than 10 times.<sup>3</sup> The first and second authors also went through a subsequent coding pass to assess the inter-rater agreement on the final set of codes. The two authors each individually coded 50 randomly selected responses (25 in-control and 25 out-of-control responses) according to the final codebook (with 25 codes); Cohen’s  $\kappa$  was 0.89, indicating almost perfect agreement. Although there were 25 codes, the inter-rater agreement was high because each code represents a well-defined feature of Twitter, and most could be directly extracted from the response without ambiguity. For example, the response “Lots of promoted tweets” was easily coded as “Ads.”

**Survey Results** As shown in Figure 1, the most frequently mentioned features were the ability to Follow and Mute (17.2%), Feed Organization (10.7%), and UI/Usability issues (10.2%). The features that participants were most likely to say left them feeling *most* in control included: Follow and Mute, Direct Messaging, Proactive Interactions, Organization Functions, Writing Tweets, Filters, and

<sup>3</sup>8 mechanisms were mentioned fewer than 10 times, including multiple accounts (7), profile pages (6), media formats (5), sharing (4), banner messages (2), reporting (2), screen time (1) and pinning tweets (1)

Search. In contrast, Bots and Spam, Ads, Topics of Content, Feed Organization, and UI/Usability issues were mentioned mostly as features that left participants feeling *least* in control. Several features including Recommendations, Privacy Concerns, Social Metrics, and Notifications received split opinions in the responses.

Table 3 provides a detailed description of the six features that were mentioned 10 times or more *and* made people feel least in control a majority of the time. In our subsequent design work, we used these six features as the basis of six challenge areas: promising domains for designers to target for improvement when designing internal and external supports to improve users’ sense of agency.

### 3.2 Design Workshop

We next conducted a 90-minute online design workshop to brainstorm a wide variety of ways to improve users’ sense of agency when using Twitter (RQ2). We created four design prompts based on the features that survey participants said influenced their sense of agency. Specifically, we asked how designers might give Twitter users more control over how they: 1) find relevant information in their feeds, 2) express themselves by writing tweets, 3) receive recommendations, and 4) receive notifications.

**Participants.** Our inclusion criteria were that participants: 1) have professional experience in a design role, and 2) use Twitter regularly. We posted a recruiting announcement on email lists and slack channels for design professionals and mid-career design students. This announcement included a link to a screener survey to collect data about demographics and professional experience, and we received 30 responses. We then selected four of these respondents based on demographic diversity and depth of design experience and invited them to participate in the workshop.<sup>4</sup> Invited respondents

<sup>4</sup>We decided not conduct additional design workshops, as designs from the first workshop provided us with a breadth of ideas to implement as internal and external supports. Other work [18, 21, 43] also recruited a similar number of designers.

**Table 3: The features that were mentioned 10 or more times in the response to the question about things that lead the users to feel most/least in control. Features are listed in the order of the frequency with which they appeared in responses to the “least in control” section of the survey. Only features that were mentioned in the “least in control” portion of the survey more often than they were mentioned in the “most in control” portion of the survey are listed here. The rest of the features can be found in the Appendix D**

Challenge Area	Description	N	Least in Control (% of responses)	Representative quote(s) (Two quotes if minority opinion on direction of control $\geq 30\%$ )
Bots and Spam	Spam, trolls, fake accounts	21	100	<i>I see lots of accounts I think I know because of the profile picture, but only to discover its their fake or cloned account. It really gets me down (P124).</i>
Ads	Ads and promoted tweets in the timeline	41	98	<i>The promoted tweets do not have a control to remove and they often have clickbait imagery to incite a response (P56).</i>
Content Topics	The authenticity, appropriateness, or emotional impact of the content the user sees	17	94	<i>I’ve come across some disturbing racist or adult material on Twitter which can be annoying. I don’t have control over what I see sometimes (P78).</i>
Feed Organization	The order of tweets and other content (e.g. users’ tweets, retweets and likes) displayed in the timeline	57	88	<i>There is an option to change the feed to the time-based view, but it automatically changes itself back to the algorithm view after some time and there is no way to opt out of this (P116).</i>
UI/Usability issues	The appearance of the interface and its interaction patterns, including unresponsiveness and navigation challenges	54	74	<i>If you click eighteen tweets deep on a thread, you can’t get back to your timeline without clicking back eighteen times (P100).</i>
Recommendations	Tweets and accounts recommended by the platform, including who to follow, explore, stories, trending topics, and “for you” section.	46	57	<i>I get annoyed by having to scroll past all of the suggestions of people that I don’t want to follow. These options come up despite the fact that I don’t want them there. (P68)</i> — <i>It is hard to keep up with all that is going on in the world. Having the trending bar makes me feel better knowing I can stay in the loop (P115).</i>

included graduate students in design-related fields and industrial designers with an average of three years of professional experience. All had used the mobile Twitter client for over a year, with two of them using Twitter daily and two using it weekly. Participants received \$40 for their time.

**Materials and Procedures.** During the workshop, the researchers first presented a summary of the survey results and then shared the four design prompts (with 15 minutes allocated to each prompt). After each prompt, participants were asked to write down their design ideas and upload their sketches to a collaborative sketchboard.<sup>5</sup> The sketchboard was divided into four sections (one for each prompt). Each participant worked independently on the shared sketchboard, and then presented their designs and rationale during a collaborative share-out session. In all, 35 design ideas were generated (see Appendix C) with accompanying sketches (see Appendix E). The research team then iteratively reviewed and refined ideas from the design workshop to define a final set of six features to implement in a mobile client for Twitter.

<sup>5</sup><https://miro.com>

**Results.** The final list of implemented designs target four of the six challenge areas raised by survey participants, including *Feed Organization*, *Content Organization*, *Recommendations*, and *Ads*. These features were categorized into two external supports, which are described in Table 4, and four internal supports, which are described in Table 5. These features allowed us to compare how internal and external mechanisms affect users’ experience.

#### 4 THE CHIRP TWITTER CLIENT

We used the design work above to create the “Chirp” Twitter client, a mobile app with two external features intended to support user agency and four internal features intended to support user agency. We did not pre-specify the number of features we would implement or the division between external and internal supports. Rather, we sought to holistically provide comprehensive support for user agency via each of these two different approaches (i.e., the external approach and the internal approach) and then to compare users’ experiences with the holistic collection of external supports that

emerged organically from our design process against users' experiences with the holistic collection of internal supports that emerged organically from our design process.

Further, we did not standardize the unit of expected impact from any one feature. We had no reason to think that each feature would affect the user experience equally, and we allowed our design process to surface both large designs (like an external support to provide a dashboard with an entire screen) and minor adjustments (like an internal support adding a small label to indicate the start of previously read tweets). Thus, the number of distinct features reflects the number of places where we made changes rather than the overall magnitude we expected these changes to have on the user's experience collectively. We strove for consistency between our internal and external approaches, and implemented the most promising features that emerged from our workshop. However, the nature of the investigation made it impossible to test all possible supports, or to ensure that our collections of external and internal supports were evenly matched. Thus, participants' responses provide useful data for comparing these two specific collections, but other supports would likely yield additional insights, and our findings should be interpreted under this limitation.

#### 4.1 The Baseline App

"Chirp" is an app for the Android mobile operating system, which we built by modifying the open-source Twitter client app, *Twidere*.<sup>6</sup> The baseline version of Chirp mirrors the official Twitter app: after logging-in with an existing Twitter account, the user sees the same feed of tweets from the people they are already following on Twitter. Because survey respondents reported that recommendations and ads influenced their sense of agency, we also wanted to include these in the baseline app for Chirp, so that we could manipulate them via internal supports. We simulated this content by sampling from trending hashtags, because the Twitter API does not provide a feed of recommendations or ads. In the baseline app, these recommended tweets were inserted into the default feed at the rate of one per every ten tweets.

#### 4.2 External Supports

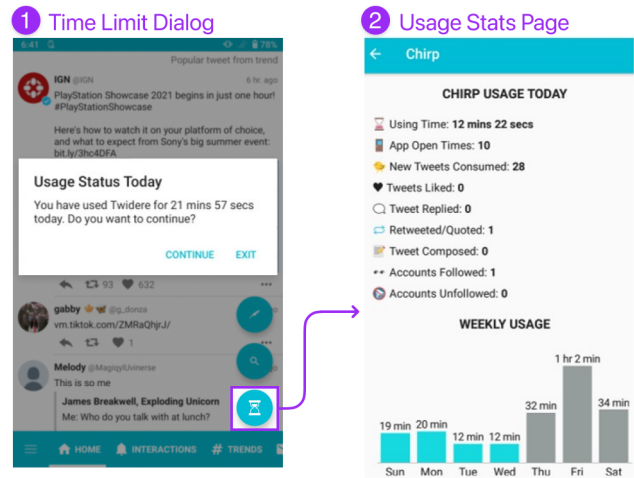
When external supports are enabled, two new features are available (see Table 4 and Fig. 2). First, a **Time Limit Dialog** pops up to notify the user about their usage time when the user has spent over 20 minutes in the app. The user can either click "exit" to exit the app, or ignore the dialog. Second, the user can click an hourglass icon on the main interface to bring up the **Usage Stats Page**. The Usage Stats Page includes metrics such as app usage time, open times, the number of new tweets viewed, tweets liked, tweets replied to, retweets or quotes, tweets composed, and accounts followed/unfollowed/muted. The usage time from the past week is also plotted in a bar chart.

#### 4.3 Internal Supports

When internal supports are enabled, four new features are available (see Table 5 and Fig. 3). First, a **Reading Progress Indicator** appears in any feed of tweets indicating that the user has already seen the tweets below. The user can manually pull the feed to refresh,

<sup>6</sup><https://github.com/TwidereProject/Twidere-Android>

## The 2 External Supports in Chirp



**Figure 2: The two external supports in Chirp: (1) Time Limit Dialog appears when the user has spent over 20 minutes in the app; and (2) Usage Stats Page shows the counts of feature use and time spent in the app. The page can be accessed from the hourglass button on the home page. Note that the feed shows recommended tweets (the first post from IGN in (1)), unless the internal support "Recommended Tweets Blocker" is enabled.**

and Chirp will automatically refresh all feeds upon launch. New tweets after the refresh are displayed above the Reading Progress Indicator. Second, a **Feed Filter** allows the user to filter tweet types by clicking a button on the top of the feed to hide replies/retweets or tweets. Third, a **Recommended Tweet Blocker** removes the recommended tweets that appear in the baseline version of the app. Fourth, Chirp hides the main feed and prompts the user to add **Custom Lists** to the bottom toolbar. The user can create and add lists by clicking the add button on the toolbar and manage list members on the list configuration page.

## 5 CHIRP FIELD DEPLOYMENT

We conducted a four-week, within-subjects field deployment with 31 participants to compare users' sense of agency across different versions of Chirp. The bundles of external and internal supports were the experimental manipulation (either turned on or off) in the different versions of the Chirp app.

### 5.1 Participants

We recruited 51 English-speaking participants from across the United States via email, online forums, and Amazon's Mechanical Turk platform. The inclusion criteria were the same as those for our formative study (see Section 3.1): owning a smartphone with the Twitter client installed and spending more than 10 minutes per day on Twitter with more than 10% of total Twitter usage occurring on the phone. Participants first completed the survey used in our formative design work and installed the Chirp client on their

**Table 4: External Supports we implemented in Chirp based on our survey and design workshop.**

Feature	Description	Example Experience (From Survey Responses)
Time Limit Dialog	Track app usage time and show a reminder when total usage time passes a threshold.	<i>I wish there was an option where I could set a time limit for myself in the app. Basically, I get sucked in and before I know it I spent way more time than I intended.</i>
Usage Stats Page	A page displaying usage metrics, including usage time, tweets viewed, tweets liked, etc.	<i>It (the mobile client) does not tell me how long I've been on or anything about my usage of Twitter.</i>

**Table 5: Internal Supports that we decided to implement in Chirp, based on our survey and design workshop. Lists is a feature that shows tweets only from Twitter users that have been added to that list, and are often setup for a particular topic or interest (e.g., sports or art)**

Feature	Description	Example Experience (From Survey Responses)
Feed Filter	A filter to control the types of tweets (original tweets, retweets, replies) shown in the feed.	<i>There is no option to just view people's own tweets, forcing you to wade through endless retweets to find what you're looking for.</i>
Reading Progress Indicator	A message saying, "You're all caught up" appears in the feed if the user has seen all new tweets.	<i>I don't want to miss any interesting posts and I can never seem to stay caught up with my feed. I end up spending way too long on the Twitter app because I do not know when to stop.</i>
Custom Lists	Collections of tweets are organized into separated feeds that are displayed on different tabs, accessible from the main screen.	<i>I create a lot of lists so that I can look for someone/specific people whose tweets I want to read on a particular topic. It keeps me focused on one issue rather than going all over the place.</i>
Recommended Tweets Blocker	Blocks recommended tweets from the default feed. Only tweets from Twitter users that the user has chosen to follow appear.	<i>Too many "recommended" feeds... There is always another tweet to look at and there may be links that take me elsewhere and before I know it an hour (or more) flies by.</i>

phones. Ultimately, after enrolling and attempting to install the app, 43 participants completed the entire four-week deployment. At the end of the deployment, we excluded participants from our analysis who used Chirp less than four days per week and less than 16 days in total. Among the removed participants, nine used Chirp less than eight days in total, three used Chirp for 10-12 days. All of the removed participants used Chirp for less than 3 days in at least one condition, and their average usage time was 5.7 minutes per day (much less than the eligible participants' 13.4 minutes per day), which meant that they might not be able to provide meaningful feedback on time-based features such as the Time Limit Dialog. This threshold was determined before beginning our analysis or examining the contents of the data that would be excluded.

Thus, the final sample was composed of 31 participants (15 men, 15 women, and one non-binary person) with an average age of 36 years ( $sd = 10.5$ ). All but one participant had used Twitter for over one year; the remaining participant had used it for over 6 months. Twenty-six participants said that, on average, they used Twitter for 5-60 minutes per day, ten for 1-2 hours per day, and seven said they used it for more than 2 hours per day. All participants (including those whose data was not analyzed) received \$10 for finishing the first week of the study, \$10 for the second week, \$20 for the third, and \$80 for the fourth (in total \$120). We also invited 11 participants to participate in a 30-minute follow-up interview, for which they received \$15 Amazon shopping gift card.

## 5.2 Four Experimental Conditions

We implemented four versions of Chirp to create all combinations of: 1) including or excluding internal supports, and 2) including or

excluding external supports. This yielded four versions of Chirp, which each served as a separate experimental condition (Table 6).

**Table 6: The four experimental conditions in the Chirp field deployment.**

Condition	External Supports	Internal Supports
Baseline		
External Only	✓	
Internal Only		✓
External and Internal	✓	✓

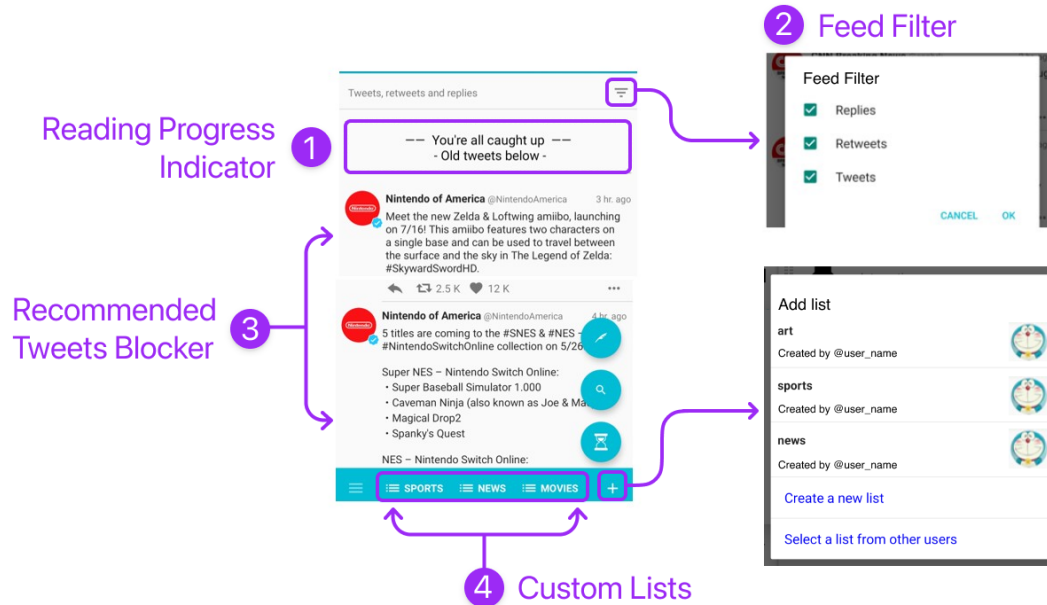
## 5.3 Data Logging & Privacy Protection

For the purpose of deployment, we implemented data logging in Chirp, including the opened times, app use time, the view time for each feed, and the view time after the reading progress indicator is shown in each feed. We also implemented support for showing pop-up questionnaires based on the experience sampling method (ESM) [12] to collect *in situ* feedback from users.

To protect participants' privacy, we transmitted the log to a Bigquery database via the Firebase platform, which supports encrypted data transmission using HTTPS. The app did not require any permissions for the data logging (i.e., all logging was performed based on in-app usage). Additionally, we deleted all logged data in Bigquery and stored it in a separate spreadsheet for analysis. The only identifiers in the spreadsheet were participant IDs.



## The 4 Internal Supports in Chirp



**Figure 3: The four internal supports in the Chirp app: (1) Reading Progress Indicator displays when all new tweets have been read; (2) Feed Filter controls which type of Tweets appear in the feed; (3) Recommended Tweets Blocker prevents recommended content from appearing the default feed; and (4) Custom Lists makes lists of specific Twitter users (e.g., Sports) easily available from the bottom bar of the home page. The user can choose which lists appear on the bottom bar.**

### 5.4 Procedure

We conducted a four-week, within-subjects deployment. All participants used each of the four versions of Chirp for one week. The order in which they viewed the four conditions was determined by a Latin square design. Each week, the version of the Chirp app on the participant’s phone would automatically change to match the assigned condition. Participants were not informed specifically about the concept of external vs. internal supports, nor were they told in advance about the features they would see each week. Each time the app changed to a new condition, the participant was sent a tutorial (in both video and pdf formats) that introduced the features in the new version. For each condition, we introduced all of the features contained in that condition, even though some of the features may have appeared in a prior week, to create a consistent experience. At the start of each condition, participants were required to open the new version of the app and take a screenshot of the new features to ensure that they viewed them at least once.

Throughout the deployment, an ESM prompt would appear after the app was used for 3 minutes and then again after every 15 minutes of use. The prompt was modeled after prior work assessing sense of agency [41] and asked the participant, “For this use of Chirp, how much did you feel in or out of control?” Response options ranged from 1 (less in control) to 5 (more in control). The ESM prompt also posed additional questions as part of a second study that is outside the scope of the current investigation; that data is not reported or analyzed here.

At the end of each week (and thus, each condition), participants filled out a survey about their experience using Chirp. The survey included questions such as, “How was your experience of [sic] using Chirp last week?” For each feature of interest that had been present that week, the survey asked the participant to provide open-ended feedback and to report how the feature made them feel by selecting from the following fixed-choice options: *More in control*, *Less in control*, *No difference*, or *Did not use it much*. The order in which the survey asked about each feature was randomized to avoid ordering effects. Full survey questions are provided in Appendix F. At the conclusion of the deployment, participants were invited to participate in a follow-up debriefing interview. We conducted 11 interviews via Zoom, where participants talked about their experience using Chirp and thoughts on each feature in detail. The interview protocol is provided in Appendix G.

## 6 RESULTS

### 6.1 App Usage by Condition

In our screening survey, six participants reported wanting to use Twitter more than they did currently, six reported wanting to use it less, and 19 wanted to maintain their current usage habits. We aggregated the amount of time each participant spent using the Chirp client in each of the four conditions (see Figure 4, left) and conducted a one-way, repeated measures ANOVA of condition on

usage time. We found no significant differences in usage time as a function of condition ( $F(3, 30) = .434, p = .73$ ).

## 6.2 Users’ Sense of Agency when Adding Internal and External Supports

To examine users’ sense of agency as a function of internal and external supports, we performed the nonparametric Aligned Rank Transform (ART) procedure [62] with the presence or absence of internal supports and the presence or absence of external supports as factors and participants’ ESM responses regarding sense of agency as our dependent measure. We found that internal supports had a small but significant effect on response scores ( $F_{(1, N=450)} = 6.73, p < .01$ ), indicating that users felt more in control when this particular set of *internal* supports was enabled. There was no significant effect of this particular set of *external* supports ( $F_{(1, N=450)} = 1.02, n.s.$ ) or the interaction of the two factors ( $F_{(1, N=450)} = 0.12, n.s.$ ). ESM scores by condition are shown in Figure 4, right.

## 6.3 Users’ Sense of Agency in Response to Each Feature

At the end of each week, we asked participants to rate the features they encountered in that week’s version of Chirp by responding with a scaled, fixed-choice response to the prompt, “*Did this feature affect the amount you felt in control of how you used Chirp?*” For each participant, we calculated an average score for each feature. Unfortunately, due to an error with data capture, we were not able to record this feedback for recommended tweets. The ratings for the remaining five features are shown in Figure 5. The *Custom Lists* feature was rated the most helpful feature in increasing users’ sense of agency, and the *Time Limit Dialog* had the lowest absolute rating.

For each participant, we averaged responses to all external supports and separately averaged responses to all internal supports, creating two measures for each participant. We ran a paired-samples *t*-test comparing these averages and found that participants were significantly more likely to say that this collection of four internal supports increased the extent to which they felt in control ( $mean = .83, sd = .25$ ) than they were to say that this collection of two external supports did ( $mean = .60, sd = .40, t(29) = 2.82, p = .009, d = .51$ ). Cohen’s *d* indicated a medium effect size.

**6.3.1 Custom Lists Increase Sense of Agency.** The most common theme participants expressed in interviews was the value they derived from the Custom Lists feature. Ten of the eleven interviewees had an unequivocally positive response to the feature, calling it, “*the greatest part about the Chirp app*” (P49) and “*a big game changer*” (P63). Participants explained that Custom Lists increased their sense of agency by helping them focus on the content they value. They said:

*“The one I enjoyed the most, I said this in the surveys, was the list grouping. I don’t think I ever had that on another app, another Twitter app, where I could categorize all of my interests like that and kind of get a condensed look at certain categories, rather than a mixture of them like on regular Twitter. I found that really helpful, because sometimes I don’t want to see the local news or something like that, or national news... It did give me*

*more control, because I could choose what I wanted to see.” (P19)*

Participants found that Custom Lists helped them easily sift through the morass of content on the platform and increased their ability to focus on the content of interest. They said things like, “*It brought me exactly where I needed and wanted to be on the app... I could be like, ‘Nope, I just want to read the news. I don’t want to see what anybody has to say’ [and] not have to follow anything meandering*” (P18). They used active language that reflected how the Custom Lists feature supported their own agency, telling us it enabled them to “*curate*” (P2), “*categorize*” (P26), “*zero in on*” (P18), and “*focus on*” (P22) content they feel is worthy of their time.

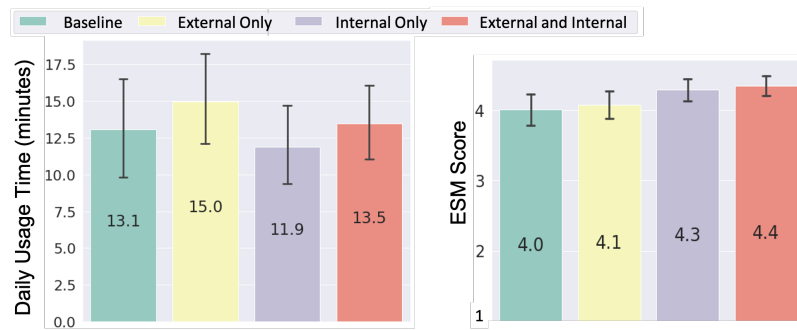
Participants described their experience in the Baseline condition as more distracting, fragmented, and overwhelming than their experience when using Custom Lists. They explained that:

*“It [Custom Lists] really helped me kind of separate modes of thinking and get people who I don’t want to hear from out of there, or reduce clutter. So that really, really helped me. And I think just that sort of being overwhelmed by information has been why I’ve always had trouble with Twitter.” (P18)*

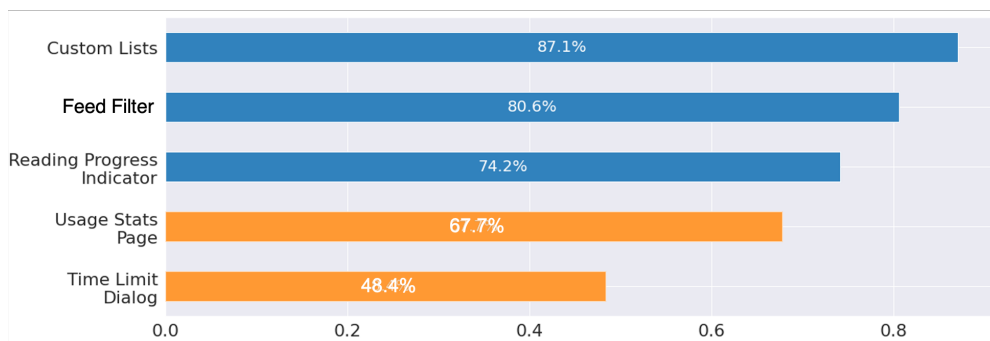
They described the default experience as chaotic and filled with irrelevant information that they had to scroll past, explaining that it requires them “*go through a bunch of nonsense*” (P8), “*has all matters mixed up together, which for me is not a super helpful way*” (P22), and forces them “*to scroll, scroll... to even find anything*” (P21). Shifting to use the Custom Lists feature to organize and view content enabled them to “*get rid of the filler*” (P26). They described the traditional Twitter by saying it entails, “*passively reading whatever it throws at you, [and] I think it [using Custom Lists] really helps with that*” (P18).

Although participants said they spent less time scrolling and sifting through the clutter when Custom Lists were enabled, they did not necessarily spend less time on the app overall. Participants said that they reduced the time they wasted with distractions and with tedious searching but increased the time they spent reading tweets they valued. People said that they instead spent more time with “*information that I actually cared about*” (P8), “*people I care about most*” (P26), and “*[the] subject I was interested in seeing*” (P22). Some people said that this led to a net decrease in usage, telling us things like, “*[Custom Lists] definitely decreased my usage just because less content to consume*” (P2) while others said that, “*[When I had Custom Lists] I’d say I was using it [Chirp] more often, for sure*” (P19). But a common sentiment led to these seemingly disparate experiences, as explained by P3: “*[Custom Lists] actually helped me personally with not using Twitter beyond the point where I can say I’m getting something out of it.*” Thus, whether it increased or decreased the quantity of time spent on the app, participants reported that it improved the *quality* of time they spent in the app. Further, users valued this increase in quality and sense of agency more than changes to the amount of time they spent in the app.

**6.3.2 The Reading Progress Indicator Increases Sense of Agency.** Similarly, interviewees reported that the Reading Progress Indicator increased their sense of agency, saying that they could more easily spend their time the way they wanted to. Eight of the



**Figure 4: Left: Average daily usage time per person by condition. Right: Average ESM response per condition. Response options ranged from 1 (less in control) to 5 (more in control). Error bars represent 95% confidence interval**



**Figure 5: For each feature, the bar plot represents the fraction of participants who said the feature increased their sense of agency. Internal supports are shown in blue, external supports in orange**

11 interviewees said that this feature consistently helped them feel more in control of their time on Chirp. For example, P18 explained that it, “*definitely helped me use [the app] in a more focused manner... it improved the quality of time. It’s hard for me to estimate [any effect on] quantity.*” Participants explained that this change arose from the simple fact that they had no wish to inadvertently spend their time re-reading content they had seen before, and that they appreciated having the interface support them in viewing new content they valued rather than leading them down an infinite feed of yesterday’s news:

*“I dislike that it [Twitter] is endless and you can kind of scroll forever, so I like that little banner that [says], ‘Hey you’ve scrolled to the point where you’ve already looked at this stuff.’ I think that helps me not look at duplicate content and go, ‘oh this, like, looks familiar’ but I’m kind of wasting my time in a sense, just viewing something that I probably looked at yesterday.” (P3)*

By proactively telling users when they were about to scroll into the content they had seen before, Chirp became less of an adversary and more of a partner in helping participants use the app the way they wanted to. Participants said that rereading old content is almost never how they want to spend their time, making it useful for social media platforms to warn users when they are about to do so, rather than seamlessly encouraging them to continue scrolling.

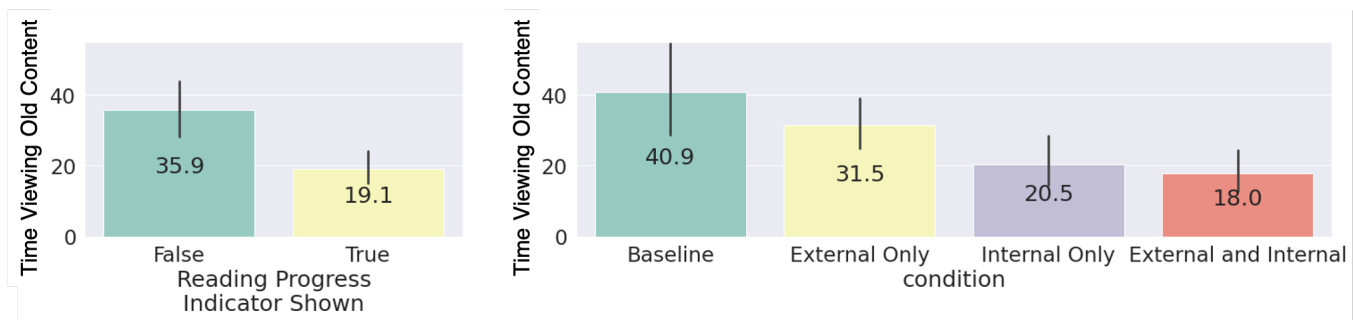
They explained that they would value this kind of interface support across all of the social platforms they use:

*“I think most people only want to go to consume new content. Most social media apps I don’t think they bother to tell you when you’ve already read something before just again because they want to keep you on the app for as long as possible... I think it’s really important to have that [the history label] in social media.” (P2)*

Nearly three-quarters of survey participants said that the Reading Progress Indicator helped them feel more in control (see Figure 5), and interviewees repeatedly told us they experienced this feature as helpful and supportive, saying things like, “*It let me know that there was nothing else to search*” (P8) and “*I thought it was pretty helpful just to know that I’ve viewed everything*” (P26).

We also examined whether the Reading Progress Indicator affected the amount of time participants spent with content they had previously seen. In each condition, we logged the amount of time participants spent viewing content they had seen before; the Reading Progress Indicator was visible and marked the transition to old content only in the conditions where internal supports were enabled. The amount of time (in minutes) that participants spent with old content is shown in Fig 6.

We log-transformed the amount of time participants spent viewing old content to make it fit gamma distributions, and we performed analyses of variance using a generalized linear mixed model



**Figure 6: The conditions with Internal Supports displayed the Reading Progress Indicator, a message saying “You’re all caught up” when the user reached the end of the new tweets in their feed. This significantly reduced the time (in minutes) that participants continued to spend viewing that feed. Error bars represent the 95% confidence interval.**

(GLMM) with a gamma link function. The independent variable was whether the Reading Progress Indicator was shown or not. The result showed that the appearance of the Reading Progress Indicator significantly reduced the time participants spent with content they had previously seen ( $\chi^2_{(1, N=2858)} = 381.42, p < .001$ ).

**6.3.3 Recommended Tweets Undermine Users’ Sense of Agency.** A plurality of participants (40%) said in their survey responses that Recommended Tweets did not affect their sense of agency. But in interviews, a majority of participants (six of eleven) described actively disliking this feature because it undermined their sense of agency. This also mirrors the split opinions regarding recommendation features that we encountered in our formative survey study (i.e., 56.5% said recommendations were one of the things that left them feeling *least* in control while 43.5% simultaneously said recommendations were one of the things that left them feeling *most* in control). For example, P2 described their experiences encountering Recommended Tweets by saying, “Most of the time I’d look past it, look at the popular tweet or popular topic and just ignore it, but sometimes I might go down that rabbit hole... most of the time discovering popular tweets is a waste of time or not meaningful to me.” P2’s impression was that Recommended Tweets usually did not affect their ability to use the app as intended, yet this content added a burden in that the participant had to work to ignore it, something P2 was often, but not always, able to do.

Other participants agreed that Recommended Tweets were an unhelpful distraction, describing them as, “a bunch of stuff that I really didn’t have a lot of interest in” (P8) and saying, “I actually didn’t really like [them]; I know what the purpose of it was, but I usually like to find my own categories and users that I follow” (P19). Participants explained that they disliked Recommended Tweets because of the lack of meaning they derived from the content, and they felt frustrated that the platform would try to entice them into spending time with content that did not hold meaning. They explained, “It kind of just felt chaotic [and] didn’t really help me get the focus I needed” (P18) and “It almost felt that you were robbing, stealing a little bit of my time” (P22).

One possible cause of participants’ negative experience with the Recommended Tweets feature could be the way it was implemented in Chirp. Due to technical and privacy limitations, Twitter does not expose personalized recommendation functions in their API. Thus,

the Recommended Tweets we displayed were selected using popular hashtags, rather than tailored to the user’s individual interests. However, in our formative survey study, participants shared both negative and positive comments about personalized recommendation features, suggesting users may appreciate having the control to decide whether to turn on recommendations.

**6.3.4 External Supports Can Inadvertently Undermine Sense of Agency.** Participants’ reactions to the two external supports (the Usage Stats Page and the Time Limit Dialog) were more ambivalent. Although many participants found things they appreciated in each of these features, others said they ignored them or even found them disruptive. For example, P19 explained that they appreciated being able to check up on their usage habits saying, “I really liked that because I could visualize my usage every day... to see what my usage habits are. I definitely liked having that.” And P2 said they were glad to have support for learning about their own behavior: “I like the idea of self-tracking. I think the usage statistics really can support that.” However, other participants were skeptical of the idea of monitoring their usage and had little interest in self-tracking. These participants explained, “I’m not going to sit up there and say, ‘Okay, I did 10 minutes today. I’ve hit that goal that I wanted to hit’ and all that... I don’t see most people wanting to see how much time they’re on it” (P8).

Participants had even less confidence that the Time Limit Dialog could be helpful, with many saying that at times it undermined their sense of agency. Six of eleven interviewees had an overwhelmingly negative experience with the feature, finding it more annoying than useful. They explained, “I think I found it more intrusive than helpful. I’d rather make that decision on my own, to close the app” (P19). Participants more often saw it as a nuisance than a source of support and described it as “just another thing I had to click to get it out of my way” (P23). P22 further explained that the very premise of an external support is inherently problematic and in conflict with internal supports:

“It’s like, ‘don’t be patronizing,’ almost. Because, don’t try to infer that because I just spent one minute more of my time in Twitter today that’s a bad thing... I separated 30 minutes of my day to check that list I created... so if you start poking me about, ‘You are spending too much time, you are spending too much time, that doesn’t make any sense, right?’” (P22)

Here, P22 points out that because the Custom Lists feature is intended to help the user engage with intention, it is contradictory to urge the user to then disengage during these moments of purposeful use. This demonstrates one way in which external supports might undermine the agency users gain through internal supports.

Participants' concern that external supports can undermine users' sense of agency was more pronounced with respect to the Time Limit Dialog than with respect to the Usage Stats Page, suggesting that not all external supports are equally helpful or harmful. To evaluate this claim, we ran a paired-samples *t*-test comparing participants' responses to the fixed-choice survey question asking how the Time Limit Dialog affected their sense control versus the Usage Stats Page. This revealed that participants were significantly more likely to say the Usage Stats Page made them feel more in control than the Time Limit Dialog ( $t(29) = -2.76, p = .005, d = -.5$ ).

**6.3.5 External Supports Are for “Other People.”** Finally, many interviewees spontaneously mentioned that they thought external supports would be more helpful for other people than for themselves. They explained that “others” might need help tracking how they spend their time or benefit from outside reminders nudging them to close their social media apps. But they went on to say that their own self-control makes such support unnecessary. For example, when discussing the Usage Stats Page, P18 said it “*probably applies more so to people who have a need to control their usage in a downward direction... so that’s not as useful for me... but I think there are people who’ve lost control of their internet usage who that would really help.*” The same participant reacted to the Time Limit Dialog by saying, “*I generally don’t think that’s as applicable to me because I don’t feel the same need to curb my internet usage as I think some people would.*”

Many other interviewees drew the same connection between the value of external supports and other people's lack of self-control. They told us things like, “*some people might have a self-control problem, where they can’t stop using it. So yeah, I think that timeout control, a timeout feature, would be nice, just to have that option*” (P19) and “*I’m a little better at the self control thing, but a lot of people have difficulty knowing when to stop scrolling*” (P3). Participants did not express this sentiment about internal supports, and no interviewees said, for example, that the Reading Progress Indicator or Custom Lists were irrelevant for people with self-control.

Participants expressed this same reaction in their weekly surveys. They reacted to Usage Stats Page saying, “*I think this would be helpful for someone trying to limit their Twitter content consumption. These are interesting stats to know, but they didn’t affect my behavior*” (P9) and to the Time Limit Dialog by saying, “*I think it is a useful feature for people that might use the app too much*” (P22). They explained that having the Time Limit Dialog would be great “*for people who get lost in their social media*” (P13).

## 7 DISCUSSION

We saw a small but significant systematic increase in users' sense of agency in response to adding a set of four internal supports that we did not see in response to adding a set of two external supports. Yet, many people still said they appreciated being able to see their usage data, suggesting there is still value in providing external supports that surface information. The time participants

spent using the Chirp app did not change systematically in response to either internal or external supports, and participants told us that adding supports improved the quality of their time with the app rather than changing the quantity. Our findings suggest several pathways for promoting user agency and digital wellbeing, such as defining experience-specific internal supports, cataloging these as abstract patterns, and moving away from time-on-task as a metric for evaluating wellbeing.

### 7.1 Using External Supports to Enhance Agency and Wellbeing

Participants' reactions to our internal supports were uniformly more positive than their reactions to our external supports. Although people's reactions will certainly vary as a function of the specific supports that are put in place, the external supports we added (a feature for monitoring usage and a timer asking them if they would like to close the app) reflect two of the most common design patterns across current digital wellbeing tools [44, 48]. Thus, participants' greater appreciation of the internal supports we created—relative to these two common features—suggests there is room for designers to improve upon the status quo. It further suggests that adding internal supports that are tailored specifically to the interface of interest is a promising approach for making such improvements.

One inherent advantage of external supports is that they are agnostic to the experiences they monitor and can be applied across an entire ecosystem of content. A time-tracking external support can operate across all apps on a smartphone, for example, and does not require participation from the individual companies creating third-party experiences. As a result, external supports may remain a useful tool for designers to deploy in instances where internal supports are not an option. We found that monitoring usage information and sharing it with the user on demand (the Usage Stats Page) was more effective in increasing agency than a dialog nudging participants to close the app (the Time Limit Dialog). Participants explained that the Time Limit Dialog at times even *undermined* their sense of agency rather than improving it, demonstrating how external supports can backfire. Thus, our results suggest that if designers continue to create standalone external supports, they will be more likely to achieve the aim of increasing user agency if they simply provide information to the user without acting on it on the user's behalf.

Prior work points out that people find this kind of passive monitoring to be ineffective in changing their behaviors in the long run, as it lacks the restrictiveness that people say they need if they are to change their habits [48]. In theory, this might suggest designing more powerful lockout mechanisms, but participants in our study were wary of intrusive features telling them what to do. Although they agreed that more restrictive interventions might be effective, they said such interventions would be better for “other people,” a stance that is consistent with people's reluctance to adopt screen time tools [59] and tendency to abandon them [34]. This past work taken together with our findings suggests the most useful external supports may be those which passively provide information and target short-term education and reflection rather than trying to encourage users to adopt tools with heavy-handed enforcement for the long term.

## 7.2 Using Internal Supports to Enhance Agency and Wellbeing

By grounding our internal supports in the specific challenges posed by the current Twitter interface, we found that we were able to move the needle slightly on people’s sense of agency in the context of consuming Twitter content. More usefully, we found that these internal supports made inroads in the specific usage scenarios they targeted, suggesting that if designers attend to the specific barriers to user agency that are most notable, they will be able to effect targeted change. People stopped scrolling through tweets they had already seen when the Reading Progress Indicator was visible, and they reported spending more time with the content they were currently interested in when their feed was segmented by the Custom Lists feature.

Collectively, people were less ambivalent about the specific internal supports we created, and they more consistently said that these features affected their sense of agency for the better. This stood in contrast to their mixed responses to the specific external supports we created, which at times they said undermined their agency and would be a better fit for someone else. These findings suggest promise in taking a tailored, platform-specific approach to designing for digital wellbeing. They further suggest the potential value of maintaining a catalogue mapping specific design patterns for internal supports to specific outcomes for users. Custom Lists could easily be adapted as a generic pattern for segmenting any space that aggregates disparate content with the potential to pull users in many different directions. The Reading Progress Indicator could be one instantiation of a pattern to annotate content to make its relevance or irrelevance more salient.

Many prior studies have shown that current screen time tools and other external supports do not yet meet users’ needs [48, 59]. Our findings suggest promise in exploring a design agenda that focuses on developing internal supports and abstracting these supports into patterns to be leveraged across experiences with similar designs. Participants pointed out that different forms of support can be in conflict, with some guiding users toward more targeted and meaningful patterns of engagement and others giving blanket nudges to cut back on all engagement. The former is more aligned with the perspectives of the participants in our study (who said they wanted to improve the quality of the time they spend on Twitter rather than the quantity) and with prior work reporting that users are seeking experiences that are inherently more meaningful, rather than tools they can use to cut off their access (e.g., [59]).

## 7.3 Usage Time as an Insufficient Metric

Participants’ descriptions of their experiences suggest that usage time has limited value as a metric for understanding their preferences. This held both with respect to creating engaging experiences and with respect to supporting agency and digital wellbeing. Many described their favorite added feature (such as the Reading Progress Indicator or Custom Lists) by saying it improved the quality of their time without changing the quantity. Others said that they could not be sure of its effect on time spent. Some people said it helped them cut back on the total time they spent on the app while others said they spent more time with Chirp once the feature was introduced, because the experience of using it became more valuable. Rather

than rereading tweets they had already seen or sifting through a sea of clutter, participants said they more easily navigated directly to the content of interest. Digital wellbeing efforts often seek to cut back on engagement, but deployment participants said that what matters most is that their engagement is meaningful.

This points to at least two conclusions. First, designers who seek to create engaging experiences cannot assume that increases in time-on-task reflect increased user interest and engagement. Digging through a frustratingly irrelevant feed of content, in some cases, increased the time participants spent but not their interest in the experience. Digging through the clutter was one of the most frequent complaints in our formative survey, and the Custom Lists feature that segmented and organized this clutter was the most appreciated feature. Helping users reduce the time they spent bushwhacking increased their engagement and appreciation for the experience. Second, designers cannot assume that decreases in time-on-task reflect improved digital wellbeing. Many participants said they used Chirp more when they had supports for increasing agency, because the experience became more worthy of their time and attention. Currently, many interventions in this space emphasize reducing time-on-task through tracking interfaces and nudges.

Although usage time is one of the most common metrics in the digital wellbeing domain, our results reveal its limitations. Instead of relying upon screen time as an outcome measure, digital wellbeing designers might develop a more nuanced understanding of user goals or adopt alternative measures such as self-reported sense of agency.

## 7.4 Limitations

This work has several limitations: 1) This work examined a small, specific bundle of internal and external supports, and there will likely be wide differences in the effectiveness of supports within either of these broad buckets. Thus, our conclusions about the comparative effectiveness of internal and external supports broadly will not apply to all possible designs or combinations of designs that fall into either category. Furthermore, we implemented four distinct internal supports in contrast to two distinct external supports, and we did not have a mechanism for predicting the magnitude of any one of these features. Thus, these results should be interpreted as providing data about our specific designs, and we cannot make strong claims about all internal supports or how they might compare to all external supports. 2) Our formative survey was conducted with participants who used Twitter more than the average use time of 10 minutes per day [53], and we may have come to different conclusions by systematically examining user groups with other habits. 3) We conducted our deployment with a small number of participants, which may have masked effects that we would see with a larger sample, potentially preventing us from seeing trends in people’s experiences with these features. 4) Our Usage Stats feature presented a static visualization. Having an interactive dashboard (such as the iOS Screen Time page) which allows the user to inspect their usage in different time granularities might be more helpful. We also did not explore other external features such as time management planning tools [42], which allow the user to plan ahead and reflect on how much time they would like to spend with a particular app.

5) Separately, prior work points out that the passive usage monitoring we explore here raises privacy concerns [61], which we did not investigate in the current study. This complicates the design agenda for external supports and suggests designers to consider the tradeoffs between supporting the user's agency over their attention and agency over their data. Future work remains to more robustly model when and how internal and external supports can benefit users.

## 8 CONCLUSION

We distilled a set of features to support user agency on Twitter by drawing on 1) a formative survey of 129 Twitter users and 2) an expert panel design workshop with four user experience designers. These target features were divided into *internal supports* and *external supports* for digital wellbeing. The *internal supports* changed aspects of Twitter's interface to reduce barriers to user agency, and the *external supports* provided standalone tools that provided users with information and nudges for self-regulating their usage habits. We implemented all of these features within a custom Twitter client we named "Chirp," and we created alternate versions that systematically enabled or disabled each set of supports. Through a four-week, within-subjects field deployment, we found that these supports increase the quality of time participants spent on Chirp without affecting quantity. These specific internal supports—but not the external ones—led to a small increase in users' sense of agency while using the app. Although participants appreciated the external support for viewing their usage data, the feature that encouraged them to close the app backfired at times and undermined their sense of agency. Many participants spontaneously concluded that heavy-handed screen-time tools of this form were best suited for "other people" who "struggle with self-control." Our findings point to a need for a digital wellbeing design agenda that promotes users' sense of agency through tailored internal supports, which target specific user scenarios and features of the experience. Cataloging these internal supports as abstract design patterns would empower designers to benefit from one another's interventions and promote digital wellbeing at scale.

## ACKNOWLEDGMENTS

We would like to thank Ningyuan Lee for his help in resolving issues during the development of the Chirp app. This work was funded in part by Facebook and National Science Foundation award #1849955. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

- [1] Eric P.S. Baumer, Phil Adams, Vera D. Khovanskaya, Tony C. Liao, Madeline E. Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, Leaving, and (Re)Lapsing: An Exploration of Facebook Non-Use Practices and Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 3257–3266.
- [2] Eric PS Baumer, Shion Guha, Emily Quan, David Mimno, and Geri K Gay. 2015. Missing photos, suffering withdrawal, or finding freedom? How experiences of social media non-use influence the likelihood of reversion. *Social Media+ Society* 1, 2 (2015), 2056305115614851.
- [3] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "I Am Definitely Manipulated, Even When I Am Aware of It. It's Ridiculous!" - Dark Patterns from the End-User Perspective. In *Designing Interactive Systems Conference 2021*. Association for Computing Machinery, New York, NY, USA, 763–776.
- [4] Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development*. sage, Thousand Oaks, CA, US.
- [5] J. Burgess and N.K. Baym. 2020. *Twitter: A Biography*. NYU Press, New York, NY, USA. <https://books.google.ca/books?id=RRQ5DwAAQBAJ>
- [6] Moira Burke and Robert E Kraut. 2016. The relationship between Facebook use and well-being depends on communication type and tie strength. *Journal of computer-mediated communication* 21, 4 (2016), 265–281.
- [7] Marta E Cecchinato, John Rooksby, Alexis Hiniker, Sean Munson, Kai Lukoff, Luigina Ciolfi, Anja Thieme, and Daniel Harrison. 2019. Designing for Digital Wellbeing: A Research & Practice Agenda. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI EA '19, Paper W17). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3290607.3298998>
- [8] Gina Masullo Chen. 2011. Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in human behavior* 27, 2 (March 2011), 755–762.
- [9] 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* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13.
- [10] Emily L. M. Collins, Anna L. Cox, Jon Bird, and Daniel Harrison. 2014. Social Networking Use and RescueTime: The Issue of Engagement. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) (UbiComp '14 Adjunct). Association for Computing Machinery, New York, NY, USA, 687–690.
- [11] Constantinos Coursaris, Michigan State University, Wietske Van Osch, Jieun Sung, Younghwa Yun, Michigan State University, LG Telecom, and Michigan State University. 2013. Disentangling twitter's adoption and use (dis)continuance: A theoretical and empirical amalgamation of uses and gratifications and diffusion of innovations. *AIS transactions on human-computer interaction* 5, 1 (March 2013), 57–83.
- [12] Mihaly Csikszentmihalyi and Reed Larson. 2014. Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology*. Springer, USA, 35–54.
- [13] Michelle Cyca. 2020. Twitter Marketing: The Complete Guide for Business. <https://blog.hootsuite.com/twitter-marketing/>
- [14] Nicole B Ellison, Jessica Vitak, Rebecca Gray, and Cliff Lampe. 2014. Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. *Journal of Computer-Mediated Communication* 19, 4 (2014), 855–870.
- [15] Christine Geeng, Savanna Yee, and Franziska Roesner. 2020. Fake News on Facebook and Twitter: Investigating How People (Don't) Investigate. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14.
- [16] Denis Gerstorf, Christina Röcke, and Margie Lachman. 2010. Antecedent-Consequent Relations of Perceived Control to Health and Social Support: Longitudinal Evidence for Between-Domain Associations Across Adulthood. *The journals of gerontology. Series B, Psychological sciences and social sciences* 66 (11 2010), 61–71. <https://doi.org/10.1093/geronb/gbq077>
- [17] 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*. Association for Computing Machinery, New York, NY, USA, 1–13.
- [18] 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*. Association for Computing Machinery, New York, NY, USA, 4746–4757.
- [19] Alexis Hiniker, Kiley Sobel, Hyewon Suh, Yi-Chen Sung, Charlotte P. Lee, and Julie A. Kientz. 2015. Texting While Parenting: How Adults Use Mobile Phones While Caring for Children at the Playground. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 727–736.
- [20] Alec R. Hosterman, Naomi R. Johnson, Ryan Stouffer, and Steven Herring. 2018. Twitter, Social Support Messages, and the #MeToo Movement. *Social media and society* 7 (2018), 69–91.
- [21] Kaori Ikematsu, Kunihiko Kato, and Yoshihiro Kawahara. 2021. LightTouch Gadgets: Extending Interactions on Capacitive Touchscreens by Converting Light Emission to Touch Inputs. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 509, 11 pages.
- [22] Philip R Johnson and S Yang. 2009. Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use. In *Communication Technology Division of the annual convention of the Association for Education in*

- Journalism and Mass Communication*. researchgate.net, Boston, MA, USA, 1.
- [23] M Laseeq Khan. 2017. Social media engagement: What motivates user participation and consumption on YouTube? *Computers in human behavior* 66 (Jan. 2017), 236–247. <https://doi.org/10.1016/j.chb.2016.09.024>
  - [24] Inyeop Kim, Hwarang Goh, Nematjon Narziev, Youngtae Noh, and Uichin Lee. 2020. Understanding User Contexts and Coping Strategies for Context-Aware Phone Distraction Management System Design. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–33.
  - [25] Inyeop Kim, Gyuwon Jung, Hayoung Jung, Minsam Ko, and Uichin Lee. 2017. Let's FOCUS: mitigating mobile phone use in college classrooms. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–29.
  - [26] Jaejeung Kim, Chiwoo Cho, and Uichin Lee. 2017. Technology supported behavior restriction for mitigating self-interruptions in multi-device environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21.
  - [27] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. Goalkeeper: Exploring interaction lockout mechanisms for regulating smartphone use. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–29.
  - [28] 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.
  - [29] Kagan Kircaburun. 2016. Effects of Gender and Personality Differences on Twitter Addiction among Turkish Undergraduates. *Journal of Education and Practice* 7, 24 (2016), 33–42.
  - [30] Kagan Kircaburun and Mark D Griffiths. 2018. Instagram addiction and the Big Five of personality: The mediating role of self-liking. *Journal of behavioral addictions* 7, 1 (March 2018), 158–170. <https://doi.org/10.1556/2006.7.2018.15>
  - [31] Minsam Ko, Seungwoo Choi, Subin Yang, Joonwon Lee, and Uichin Lee. 2015. FamLync: Facilitating Participatory Parental Mediation of Adolescents' Smartphone Use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Osaka, Japan) (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 867–878.
  - [32] 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* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 998–1010.
  - [33] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: A Group-Based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (Vancouver, BC, Canada) (CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1235–1245.
  - [34] Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2018. Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 95 (nov 2018), 25 pages.
  - [35] 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 (Yokohama, Japan) (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 229, 14 pages.
  - [36] M. Lachman and S. L. Weaver. 1998. The sense of control as a moderator of social class differences in health and well-being. *Journal of personality and social psychology* 74 3 (1998), 763–73.
  - [37] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* 1, 1 (1977), 159–174.
  - [38] Heyoung Lee, Heejune Ahn, Samwook Choi, and Wanbok Choi. 2014. The SAMS: Smartphone addiction management system and verification. *Journal of medical systems* 38, 1 (2014), 1–10.
  - [39] 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 (Luleå, Sweden) (MUM '13)*. Association for Computing Machinery, New York, NY, USA, Article 43, 2 pages.
  - [40] Edwin A Locke. 2002. Setting goals for life and happiness. *Handbook of positive psychology* 522 (2002), 299–312.
  - [41] 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, Article 368, 17 pages.
  - [42] John R. Lund and Jason Wiese. 2021. Less is More: Exploring Support for Time Management Planning. In *Designing Interactive Systems Conference 2021 (Virtual Event, USA) (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 392–405.
  - [43] Yuhan Luo, Peiyi Liu, and Eun Kyoung Choe. 2019. Co-Designing Food Trackers with Dietitians: Identifying Design Opportunities for Food Tracker Customization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13.
  - [44] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleef, 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 (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–18.
  - [45] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotko, Jun Zhao, Max Van Kleef, 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.
  - [46] Janet Metcalfe and Matthew Jason Greene. 2007. Metacognition of agency. *Journal of Experimental Psychology: General* 136, 2 (2007), 184.
  - [47] Ally Mintzer. 2020. Paying Attention: The Attention Economy. *Berkeley Economics Review* 1, 1 (Mar 2020), 1.
  - [48] 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 (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14.
  - [49] James W. Moore. 2016. What Is the Sense of Agency and Why Does it Matter? *Frontiers in Psychology* 7 (2016), 1.
  - [50] Meredith Ringel Morris. 2014. Social Networking Site Use by Mothers of Young Children. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (Baltimore, Maryland, USA) (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1272–1282.
  - [51] Martha C. Nussbaum. 2013. *Creating capabilities: the human development approach*. Belknap Press of Harvard University Press, Cambridge, MA.
  - [52] 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 (Barcelona, Spain) (MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages.
  - [53] Mark Peterson. 2021. How Much Time Do People Spend On Social Media In 2021? <https://famemass.com/time-spent-on-social-media/>. FameMass.
  - [54] Laura Portwood-Stacer. 2013. Media refusal and conspicuous non-consumption: The performative and political dimensions of Facebook abstention. *New Media & Society* 15, 7 (2013), 1041–1057.
  - [55] Anne Elizabeth Sanders and AJ Spencer. 2005. Why do poor adults rate their oral health poorly? *Australian dental journal* 50, 3 (2005), 161–167.
  - [56] 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 (Toronto, Ontario, Canada) (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 773–782.
  - [57] Stefan Stieglitz, Deborah Bunker, Milad Mirbabaie, and Christian Ehnis. 2018. Sense-making in social media during extreme events. *Journal of Contingencies and Crisis Management* 26, 1 (2018), 4–15.
  - [58] 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 (2017), 1552.
  - [59] Jonathan A. Tran, Katie S. Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14.
  - [60] Mariek M P Vanden Abeele. 2020. Digital Wellbeing as a Dynamic Construct. *Communication Theory* 31, 4 (10 2020), 932–955.
  - [61] Kelly Widdicks. 2020. When the Good Turns Ugly: Speculating Next Steps for Digital Wellbeing Tools. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. Association for Computing Machinery, New York, NY, USA, Article 89, 6 pages.
  - [62] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 143–146.



## A TWITTER USAGE SURVEY

### A.1 Section 1: Eligibility Survey

- What is your age?
- What is your gender identity?
- What types of smartphone or tablet system do you use? (iOS/Android/other)
- Is the Twitter mobile app installed on your phone or tablet?
- When you use Twitter, about what percent of your time do you use it on your phone or tablet?
- In the past week, on average, approximately how much time PER DAY have you spent actively using Twitter overall?

### A.2 Section 2: Experience Survey

- What are 3 things about the mobile app that lead you to feel MOST IN CONTROL over how you spend your time on Twitter?
  - Thing 1/2/3 in the mobile app - please describe.
  - How does Thing 1/2/3 make you feel MOST IN CONTROL over how you spend your time on Twitter?
- What are 3 things about the mobile app that lead you to feel LEAST IN CONTROL over how you spend your time on Twitter?
  - Thing 1/2/3 in the mobile app - please describe.
  - How does Thing 1/2/3 make you feel LEAST IN CONTROL over how you spend your time on Twitter?

## B FIRST ROUND CODEBOOK

The 33 features (codes) emerged from the first round coding were: *Notification, Writing, Character-limit, Layout/UI, Twitter Metrics (followers, etc.), Follow/mute, Filter, Memory of Usage (maintaining history of the twitter usage such as the position of the timeline), List, Bookmark, Privacy, Profile Page, Recommendations, Search, Share, Schedule Tweets, Social Interaction (like/retweet/reply), Ads, Bots/fake accounts, Censorship, Addiction, Topic of the content, Feeds Organization, Privacy Settings, Misinformation, Media Format, Usability Issues, Multiple Accounts, Screen Time, Reporting Functions, Pin Tweet, Connect Phone Contacts, Banner Messages*

## C DESIGN IDEAS FROM THE DESIGN PANEL

### C.1 Curating Tweets and Finding Relevant Information

- (1) Adding more granularity to filter/search function, such as by location, account, language, with/without media, regex expression
- (2) More ways to sort the replies, e.g. by time, by responses, by likes
- (3) Unless I click "all reply", or only show the original tweets in the main feeds
- (4) More organized way to see tweets - humor vs. professional information
- (5) Based upon who you follow/tweets you like, recommended accounts to follow
- (6) Separate other users' liked tweets
- (7) Tag different users: "important" vs. "memes" etc.
- (8) One button to turn off all suggestions from my list

- (9) Sort the tweets by newest/trending (my current feed is not chronological)

### C.2 Expressing Oneself

- (1) Tag your own tweets into categories
- (2) Auto destroy tweets in a certain time
- (3) Letting the users decide by themselves that who can see their tweets
- (4) "Quick reply" provides more memes for the users to express their attitude
- (5) More features to interact with people - currently just a poll option and liking/retweeting
- (6) Sound/audio tweets
- (7) Ability to hide content from feed, display only user
- (8) "Visible only to me" option when tweeting, and the option to make it public in the future
- (9) Schedule tweets: edit it now, release it later
- (10) Undo tweets (in 30s)

### C.3 Recommendations

- (1) A "StumbleUpon" experience for Twitter, where you get a window into what a different universe on Twitter looks like
- (2) Ability to "uninterested" the recommendations
- (3) Organize the "for you" by content categories, e.g. tweets, topics, people
- (4) My notifications are always full of "news for you" which I pretty much don't care about
- (5) Recommendations integrated into the timeline based upon certain tweets or likes
- (6) Recommendations on different platform types
- (7) Have one centralized place for recommendations. I remember seeing recommendations in other users' account page
- (8) Ability to select don't recommend this user or topic again
- (9) Recommend who to unfollow
- (10) Choose how "wild" the algorithm recommends you stuffs

### C.4 Notifications

- (1) Add summary of notifications (like daily digest)
- (2) customize notifications (when, who, what)
- (3) Organize the "for you" by content categories, e.g. tweets, topics, people
- (4) Within messages, you'll get a notification someone reacted to a message but not know what message it was for
- (5) better algorithm for "x user just tweeted after a while" (see Instagram for users posting on story)
- (6) Develop a widget for iPhone and Android: a notification board. Will update without notifying you (background push)

## D DESIGN MECHANISMS

**Table 7: The design mechanisms that were mentioned 10 or more times in the response to the question about three things that lead the users to feel most/least in control. Design mechanisms are shown in order of frequency in response to feeling "least in control." This table contains the ones with less than 50% least in control portion. The first part can be found in Table 3**

Design mechanism	Description	Response count	Least in control (% of the responses)	Representative quote(s) (2 quotes if minority opinion on direction of control $\geq 30\%$ )
Privacy concerns	Privacy related settings and disclosures on how the data is used	33	48.5	<i>I am concerned about how much info Twitter has about me that they are using to target ads towards me. (P69)</i> — <i>Being able to set my account to private is a huge help. This makes makes me not need to worry about strangers seeing my posts. (P22)</i>
Social metrics	Metrics of Twitter, including follower counts, number of likes	11	45.5	<i>I have no idea of who unfollows me unless I go through all of my followers. (P22)</i> — <i>I like that it keeps track of how many people follow me and who I follow back. (P104)</i>
Notification	System and in-app alerts when there are new tweets, interactions such as replies, recommendations, etc.	50	36.0	<i>Sometimes I'll get an invasive notification just to let me know that 12 hours ago someone made a popular tweet on a topic I don't care about. (P85)</i> — <i>I feel alerts help me in control of my time as I can scroll quickly through the alerts to see what is posted and click on it if I want to further explore. (P93)</i>
Search	Searching tweets, accounts and topics	31	25.8	<i>Searching for topics is easy and I am able to sort it by top or latest post.</i>
Filters	Blocking keywords and topics	16	25.0	<i>I don't want to see things about sports or things I hate. Filtering gives me control over what I view. (P123)</i>
Writing tweets	Composition of a tweet	22	22.7	<i>I post special things for my followers to keep up with me and share my interests with them. I'm in charge of this show. (P96)</i>
Organization function	Lists, bookmarks, and functions that help organizing tweets	10	10.0	<i>I create a lot of lists so that I can look for someone whose tweets I want to read on a particular topic. (P126)</i>
Follow and mute function	Following or muting accounts	92	2.2	<i>The ability to mute accounts lets me have an enjoyable time by getting rid of annoying users in my feed. (P110)</i>
Proactive interactions	Like, retweet, reply	21	0	<i>I often retweet things I find important and I like a bunch of tweets, which makes me feel engaged with the community. (P129)</i>
Direct messaging	Messaging with other users and groups	11	0	<i>I can directly message people or someone if I need help with something or share my opinions. (P112)</i>

## E DESIGN SKETCHES FROM THE DESIGN PANEL

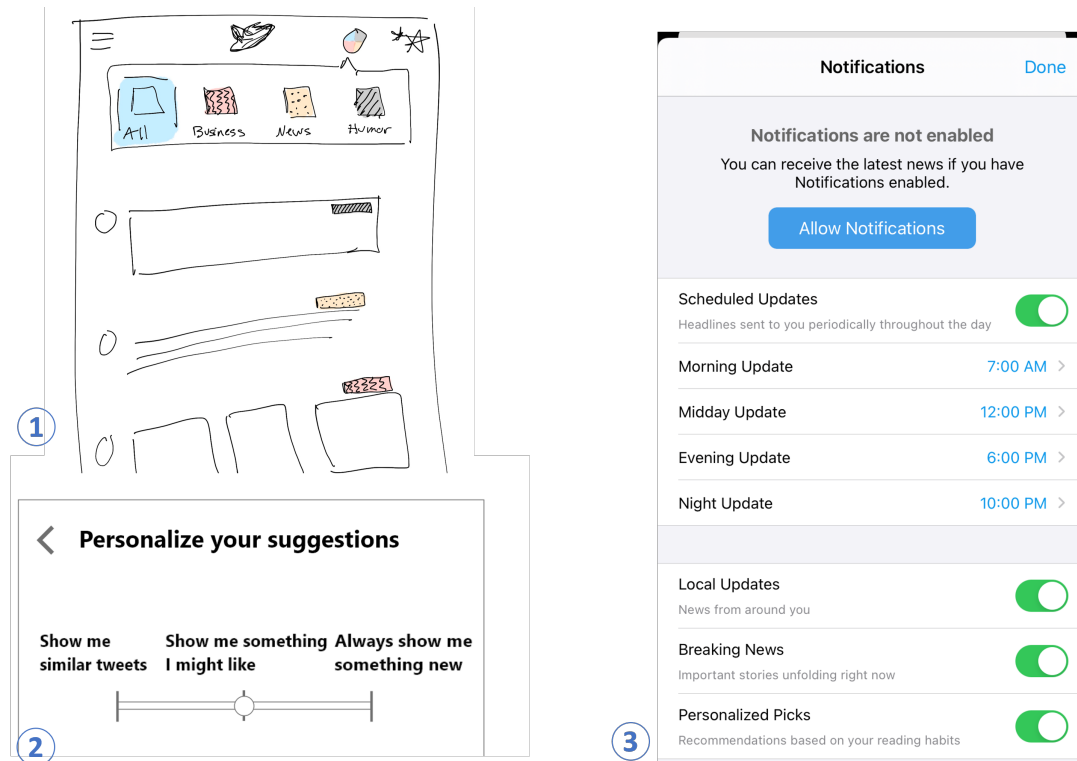


Figure 7: Sketches from the expert design panel: 1) a timeline that is organized by topics, such as art, business, or news, which can be shown or hidden; 2) adjusting the personalization algorithm: the user can adjust the novelty of the recommendation algorithm from *showing only similar tweets* to *always showing new tweets*; 3) advanced notification control: one participant recommended the notification setting page of the *News Break* app, which allows the user to restrict notifications to certain times of the day

## F WEEKLY DEPLOYMENT SURVEY

- How was your experience of using Chirp last week?
- For the last week, how much did you feel in control of your Chirp usage? Why do you feel that way?
- How satisfied were you using Chirp last week?
- (For each feature of that week, display a screenshot of the feature, and then ask)
  - How do you like or dislike about this feature?
  - Did this feature affect the amount you felt in control of how you used Chirp? (More in control/No difference/Less in control/I did not use it enough)
- Any comment for this study or your experience using Chirp?

## G DEPLOYMENT STUDY INTERVIEW SCRIPT

- A general question: Reflect on the features, how do they affect your daily tweet consumption?
- (For each feature in the study, display a screenshot of the feature, and then ask)
  - How do you like or dislike about this feature?
  - Do you think you used Chirp more, less, or about the same amount because of this feature?
  - If you could add this feature to Twitter would you want to do that or would you rather leave Twitter the way it is?
- Is there anything you would like to change about the way you use Twitter in the future?
  - What influenced your thoughts around this?
- Do you have any improvement suggestions, or features that you would like to have on Chirp?
- Now that the study has completed, would you like me to help you delete Chirp from your phone?