# Usability Hasn't Peaked: Exploring How Expressive Design Overcomes the Usability Plateau

Frank Bentley
Google
San Francisco, California, USA
fbentley@google.com

Lennard Lukas Schmidt
Google
Munich, Germany
lennardschmidt@google.com

Alyssa Sheehan
Ipsos
Atlanta, Georgia, USA
alyssa.sheehan@ipsos.com

Bianca Gallardo
Ipsos
Chicago, Illinois, USA
bianca.gallardo@ipsos.com

Ying Wang
Ipsos
Chicago, Illinois, USA
ying.wang3@ipsos.com

## Abstract

Critics have argued that mobile usability has largely been optimized, and that only incremental gains are possible. We set out to explore if the newest generation of design systems, which promote greater flexibility and a return to design basics, could produce substantially more usable designs while maintaining or increasing aesthetic judgments. Through a study with 48 diverse participants completing tasks in 10 different applications, we found that in designs created following Material 3 Expressive guidelines, users fixated on the correct screen element for a task 33% faster, completed tasks 20% faster, and rated experiences more positively compared to versions designed using the previous Material design system. These improvements in performance and aesthetic ratings challenge the premise of a usability plateau and show that mobile usability has not peaked. We illustrate specific opportunities to make mobile experiences more usable by returning to design fundamentals while highlighting risks of added flexibility.

## CCS Concepts

• **Human-centered computing → Empirical studies in interaction design**; **User interface toolkits**; **Graphical user interfaces**; **Usability testing**.

## Keywords

Design Systems, Expressive Design, Usability, Eye Tracking, Mobile Computing

**ACM Reference Format:**
Frank Bentley, Lennard Lukas Schmidt, Alyssa Sheehan, Bianca Gallardo, and Ying Wang. 2026. Usability Hasn't Peaked: Exploring How Expressive Design Overcomes the Usability Plateau. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3772318.3790373


*CHI '26, Barcelona, Spain*

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790373

## 1 Introduction

Creating user interfaces that are more usable has been a goal of the HCI community since its inception. Early research in the field, especially in the first wave of HCI [9], focused heavily on task completion times and understanding how the spatial organization of a user interface could lead to more efficient use. Principles such as Fitts' Law [23] emerged from this work, which led to recommendations for target sizes, such as the size of a button or menu.

Over the years, this type of guidance evolved into corporate style guides and eventually design systems, such as Google's Material Design [28] or Apple's Human Interface Guidelines [4]. These guidelines have evolved through multiple iterations, as seen in Figure 1, to meet changing aesthetic taste and in response to large quantities of usability studies run by design system teams to better optimize for usability and accessibility.

Material Design is the design system for all Google products, as well as the default open source design system for any app created by developers for the Android platform, the most widely used mobile operating system [55]. Billions of people regularly experience designs created with this system. The latest version of the Material Design system, Material 3 Expressive [28], has taken a different approach to systematizing design, moving away from single recommended component variants to offering increased flexibility. The guidelines encourage designers to return to the basics of interface design by more fully considering the use of size, shape, color, typography, motion, and hierarchy to create more usable interfaces. This approach empowers designers to move beyond a one-size-fits-all model and craft user experiences that are not only functional but also contextually appropriate in an increasingly complex digital landscape.

Material 3 Expressive includes updates across components. For example, new Button variants, such as Extra Large and Extra Small, are available to guide a user's attention to a key action, in contrast to previous versions of Material Design where every button had to be 48 density-independent pixels (dp) tall regardless of its use [27]. Lists, Menus, and Toolbars now use color-based containment to better group like items together in contrast to previous versions that built everything from a white background. Since Material Design is the default design system for any app created for the Android platform, it is critical to explore how changes to this open source system impact user perception and behavior given the widespread use of Android globally [55].

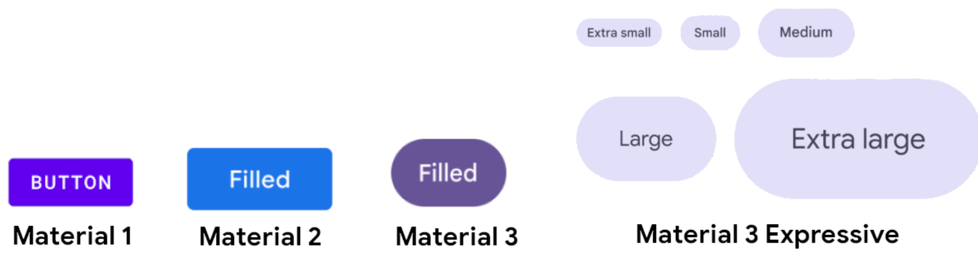


**Figure 1: The evolution of buttons in the Material Design system over the past 11 years.**

We wanted to know to what extent interfaces produced with these new guidelines and components could achieve greater usability while maintaining or exceeding aesthetic preference across a diverse sample of users. Specifically, we had the following research questions:

- RQ1: Can designs produced following Material 3 Expressive guidelines be more usable than the previous design system in terms of time to fixate and time to tap on both primary and secondary UI elements on the screen?
- RQ2: Are Material 3 Expressive designs preferred and seen as more modern and desirable?

To answer these questions, we ran an in-lab experiment with 48 diverse participants interacting with design updates for ten widely-used applications that were created following these new Expressive guidelines. Participants used eyetracking glasses and completed tasks in each application in a randomized order covering both Expressive designs and non-Expressive Material 3 baselines that matched their currently available designs. We found that Expressive designs led to 33% faster fixation times on the correct UI element for each task as well as 20% faster overall task completion times. We saw even larger effects for participants over 45 years of age. Participants also significantly preferred the Expressive designs on a variety of aesthetic attributes.

Our findings show that mobile usability has not peaked, and that there is still a large untapped opportunity to make common mobile experiences more usable for all by returning to basic design fundamentals and increasing flexibility in size and color use. This runs counter to recent claims by Jakob Nielsen that *"[t]he low-hanging fruit of terrible usability has long since been picked,"* and that there is a low return on UX work when it comes to usability [42]. Others similarly argue that UX Research as a discipline, and usability testing in particular, are now less important [21, 25]. In a landscape of high reliance on mobile devices, with average global screen time for 16-64 year olds reaching 6 hours and 40 minutes a day [61], we argue that the need to reinforce foundational usability in design is more pressing than ever. Specifically, focusing on improving task completion times can create large efficiency gains over the thousands of interactions a user makes during those 6 hours and 40 minutes.

While design systems like previous versions of Material Design have standardized UI patterns, their rigidity may have inadvertently created a usability plateau. This study investigates whether a return to fundamental design principles—flexibility in color, size, and hierarchy—as embodied by Material 3 Expressive, can overcome this plateau.

## 2 Background

We will begin by reviewing the literature and background on a few essential topics. First, we will explore the history of design systems to understand what led to the largely prescriptive design systems of today. Then, we will explore ways of measuring usability in the HCI literature as well as critiques on the state of mobile usability and recent studies using eye tracking as a method to measure usability.

### 2.1 Design Systems

Complex systems that have multiple user touch points [53] can be confusing to navigate without central design principles. One of the first instances of what we might call a modern design system came from Frank Pick of the London Underground starting in 1906 [24]. He noted the inconsistencies of signage and navigation across the transit network and the need for common wayfinding, fonts, icons, and maps to make the system more usable. The resulting standards provided a consistent experience for transit users from one station to another. Similar to modern design systems, the goal was to increase usage across the entire transit system by providing common visuals across interaction points leading to an experience that was easier to use and more predictable.

The NASA Graphics Standards Manual [1], published in 1976, was another example of a set of standard practices to be used across the agency. Not only did this manual cover how and where to use the new NASA "worm" logo, but it covered various grid layouts for NASA publications, a color system with contrast rules, and design guidelines for standardized physical signs and vehicle markings.

The need for standardization became paramount as interactions moved from the physical to the digital realms. Companies adopted various style guides to ensure standardization across their products. Perhaps the most influential of these in early years was the Apple Human Interface Guidelines (HIG), published in 1987 for "The Apple Desktop Interface" [5]. These guidelines not only presented common user interface components such as buttons, checkboxes, and dialogs but also discussed how to best use them. Since the introduction of the HIG, the guidelines have been continuously

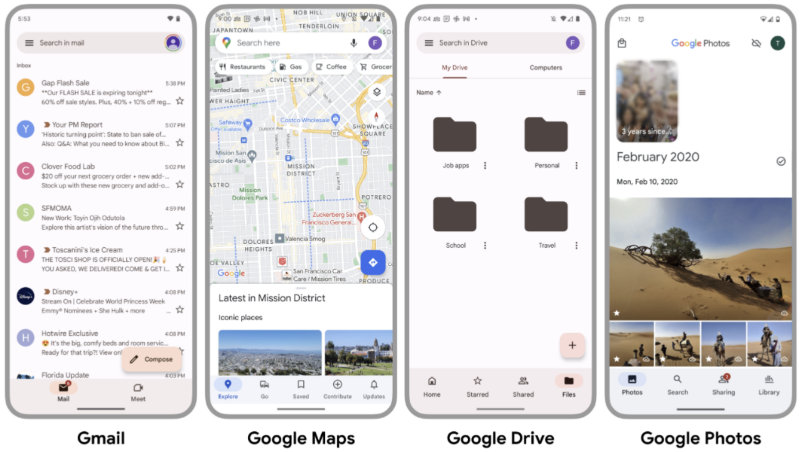


**Figure 2: A variety of apps all built with Material Design 3 as they existed in 2023, which tended to look the same and lack unique feeling for different types of experiences.**

updated to its current form [4], which now encompasses interaction on a variety of device types.

The user interface design landscape of the early 2010s was characterized by a stark dichotomy, a pendulum swinging between two opposing philosophies: the rich, tactile world of skeuomorphism and the stark, minimalist plane of flat design [11]. This schism created significant challenges in usability, consistency, and development efficiency, setting the stage for a unifying synthesis. In parallel, the shift of usage from desktop to mobile created the need for common mobile design patterns, leading Google to release the open source Material Design system in 2014 [26]. Inspired by overlapping sheets of paper, this system popularized the term “Design System” and became the basis for not only Google products but also the default design system for any app created for Android devices. A variety of studies have shown that following previous versions of Material Design guidelines have led to increased usability [6, 19, 22, 44, 46, 57]. There are no known studies prior to ours on the latest version of Material 3 Expressive.

From the early days of mobile and desktop interfaces, design systems have matured. The visual styles of components have changed with the tastes of each era as well as with advancing technological capabilities. We can trace this evolution from the original stroke outline button of the first Mac, to skeuomorphic designs of early smartphones, to more minimalist designs of the mid 2010s, to Expressive or semi-transparent designs of today [59]. However, most design systems have maintained a relatively small set of variants for basic components. Tap targets for Android Buttons were 48 dp [27], iOS ones 44 pt [3], and there was little room for variance if one wanted to remain compliant to the design system guidelines. While this provided standardization, it did not give designers the flexibility to create differentiated designs for different use cases, and mobile design started to feel the same, as seen in Figure 2.

Material 3 Expressive, released in May 2025, prompted designers to go back to basics and explore variations in size, shape, color, containment, type, and motion to design interfaces that were fit to the feeling that the designer wanted to evoke [28]. This opened up the ability for designers to go beyond constrained and equally sized components and make interfaces that are optimized for the situation. Enabling flexibility empowers designers to move beyond a one-size-fits-all model and craft user experiences that are functional and contextually appropriate in an increasingly complex digital landscape.

Our study, the first on the application of Material 3 Expressive, demonstrates the ability of more flexible design systems to deliver higher user satisfaction through more usable interfaces.

## 2.2 Usability

Usability has been a focus of the HCI community from the beginning. In 1983, Card et al. [14] explored the idea of a model human processor that would help designers predict how long a user would take to perform a task. Fundamental work continued throughout the 1980s. Schneiderman created his “Eight Golden Rules of Interface Design” [52] in 1985 which promoted consistency in an interface in terms of color, layout, and typography to enhance usability. Nielsen's heuristics for usability [39] and the previously mentioned 1987 edition of the Apple Human Interface Guidelines [5], which focused on designing for usability, continued work towards standardized principles to guide digital designers. Much of this first wave of HCI was focused on usability, with key metrics such as task completion time and error rates seen as critical to understanding and evaluating the usability of a system [9]. Since then, a greater focus on the broader User Experience (UX) has occurred, which has accepted a more encompassing definition of UX covering everything from "measurable aspects of UX such as psychological responses and user task performance (cf traditional usability metrics)” as well as “non-utilitarian” or “hedonic” aspects such as “user affect” [35].

The Nielsen Norman Group currently defines usability as a set of five factors: Learnability, Efficiency, Memorability, Errors, and Satisfaction [41]. While there are many established methods to measure usability, by combining eye tracking fixation data and task completion time, one can measure both learnability and efficiency.

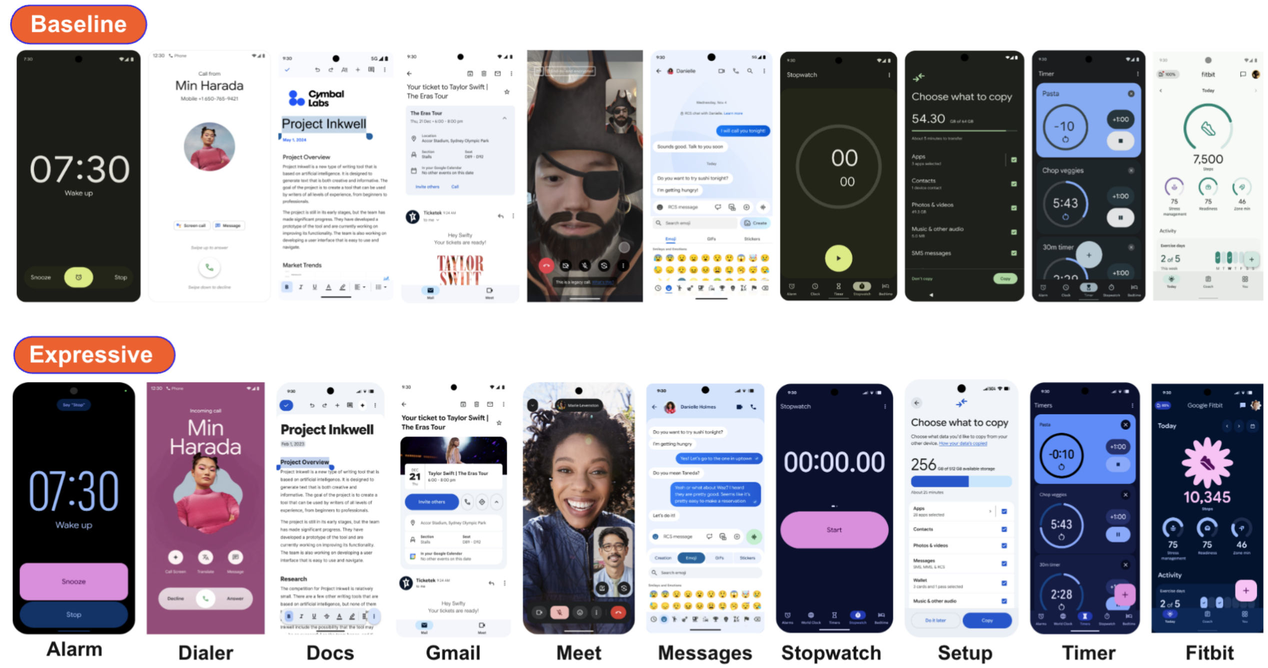


**Figure 3: The 10 application designs tested in the study, each in Baseline and Expressive conditions. These 20 screens were displayed in a random order for each participant**

Satisfaction can be measured through questions around preference and overall ease of use. Many studies over the years have used these types of metrics as a proxy for usability [17, 32, 60, 62, 63] as eyetracking has become an established tool in the HCI literature and UX Research practice to measure usability [8, 10, 20, 30, 37, 47]. In addition, more hedonic attributes such as "fun" can be measured to understand more experiential aspects of the User Experience [35].

Usability in general has been improving over time as companies realized the monetary benefits of efficient interfaces as well as the ability to reach broader audiences through increased accessibility. Companies large and small have research teams devoted to increasing the usability of their products [34, 36, 56]. One might wonder if we are reaching diminishing returns on future usability work, as Nielsen strongly argued in a recent post [42]. We questioned if this was the case, and set out to run this study to see if a more dramatic change to a design system, allowing for much greater variations in size, shape, color, and containment could yield much larger usability gains over what might just be a local optimum with the previous generation of constrained design systems.

## 2.3 Method

We conducted an in-person study in Chicago where participants could interact with both Expressive and current versions of designs for ten diverse applications. We tested a wide range of applications, including productivity, utility, fitness, chat, and video apps, to best understand the impact of these design principles across wide categories of products (see Figure 3). 48 participants completed the study, representing a broad sample of Chicagoans, ranging in age from 18 to 62 (33% 18-24, 33% 25-44, 33% 45+), from diverse ethnic backgrounds (40% White, 27% Black, 19% Hispanic, 12% Asian, 2% Native American), and gender-balanced (48% women, 48% men, 4% non-binary).

To generate the screens for testing, we enlisted existing product designers for ten diverse products to apply the principles of Material 3 Expressive to their own Android applications. All of the applications used the previous version of Material Design at the time of the study, which created a consistent baseline condition of current designs. The designers were given a Figma Design Kit with the latest Material 3 Expressive components as well as access to the guidelines for how to design with them, but were not given any specific instructions to focus on usability. They had several months to complete their designs and in the end we received baseline (current Material 3) and Expressive designs for Alarm, Dialer, Docs, Fitbit, Gmail, Meet, Messages, Android Setup, Stopwatch, and Timer, as shown in Figure 3. After receiving the designs, researchers and designers reviewed them for compliance with the components and guidelines of Material 3 Expressive, ensuring designs properly adopted the principles. We then created tasks for each design, as shown in Table 1. These task descriptions were carefully created to not be leading (for example by not using labels present in the interface). Tasks were a mix of primary (e.g. answer an incoming call) and secondary (e.g. send a voice message) actions on the screen, as we wanted to ensure that Expressive design did not just affect the usability of the main or largest action on the screen. The baseline applications were chosen to be existing, highly usable applications that represent decades of work of highly skilled Design and UX Research teams. These are exactly the types of interfaces that Nielsen thought would be the hardest to show large usability gains [42].

Participants engaged with both sets of designs using a mobile test application that was specifically built for our study. The test app included these 20 screens (10 expressive and 10 current), as shown in Figure 3, and was installed onto an Android device used by all participants in our study. Screens were presented in a random order for each participant, with a screen in between tasks that displayed only the name of the next random app. The moderator would read the next task based on the app name that was shown. The phone screen would then turn black for 5 seconds to reset the user's gaze before showing the screen for the user to interact with. The study was conducted by researchers not affiliated with the company that made the designs.

To answer RQ1 above, we created two testable hypotheses:

(1) H1: Participants will notice the UI component needed to complete the task faster in the Expressive conditions
(2) H2: Participants will tap the UI component needed to complete the task faster in the Expressive conditions

These are complementary metrics as one might notice an element quickly but not know if it is the right one to tap. Or they might not see a UI element right away, but tap it the second they see it. To capture these measurements, participants wore Tobii Pro Glasses 3 eye tracking glasses [58], and Tobii's software was used to capture the time for the participant's eyes to first fixate on the correct UI component for the task. In addition, our study application also captured the number of milliseconds from when the screen was first displayed until the participant tapped the correct UI element.

After completing tasks in all 20 screens, participants were shown each screen one by one and asked to rate it on a five-point scale on five attributes related to visual aesthetics: Appealing, Clean, Fun, Modern, and Easiness. Together these capture the satisfaction aspect of usability. Specifically, Appealing captures satisfaction with the visual aspects and Easiness captures the perception of overall usability. Modern and Clean relate to aspects of the visual design shown to be correlated with use. And Fun aimed to capture the emotional aspects of design that Material 3 Expressive claims to achieve, and is related to the overall usability goal of being satisfied with an interaction. Perrig et al. [45] have explored how aesthetic attributes such as these are correlated with user performance (usability) and subjective experience with a design.

Participants were then shown the Expressive designs and Baseline designs together and asked which designs they preferred for each app and why. These responses formed the basis of our qualitative data set. Each session lasted approximately 45 minutes and participants were compensated for their time. Each session was recorded and transcribed, and adhered to all procedures for studies with human subjects at our institutions.

**Table 1: Tasks as they were read to users before each task.**

| App | Task |
|---|---|
| Alarm | Imagine that your morning alarm is going off and you want to pause it for 10 minutes until it goes off again. Please show me how you would do this. |
| Dialer | Imagine that your friend Min is calling you. Go ahead and pick up the call. |
| Docs | Imagine that you are reviewing a document on your phone. You want to unbold the text at the current position. Please show me how you would do so. |
| Fitbit | Imagine that you just finished a workout and would like to manually log your activity on your Fitbit app. Show me how you would log a new activity there. |
| Gmail | Imagine that you bought tickets to Taylor Swift's Tour and have opened the confirmation message in your email app. You want to share the ticket details with a friend who is attending with you. Go ahead and do this. |
| Meet | Imagine that you are on a video call with your friend. You realize that you are on mute and would like to unmute. Please show how you would do that. |
| Android Messages | Imagine that you are texting your friend and would like to send an audio message. Please show how you would send an audio message. |
| Setup | Imagine you are setting up a new phone and you need to select which data to copy over to your new device. You have selected all that you wanted to copy over and want to continue. Please show how you would get to the subsequent screen of the setup process, accepting all of the options on the screen. |
| Stopwatch | Imagine that you want to use the stopwatch on your phone to time how fast your friend can sprint. Your friend is about to go. Tap the appropriate part of the screen to get the timer going. |
| Timer | Imagine that you are doing laundry and you want to set a timer on your phone to remind you to check laundry. Please show me how you would create a new timer on your phone. |

## 2.4 Data Analysis

To develop a comprehensive understanding of user interaction with Expressive designs, our study employed a mixed-methods approach, integrating both quantitative performance metrics and qualitative insights. Quantitative data was captured and analyzed with Tobii eye tracking glasses and Tobii's software provided time to first fixation on the region of interest for each screen that matched the tap target to complete the task. This usability performance data was then joined with demographic information and participant sentiment scores. A fixed-effects linear regression model [2], commonly used in HCI papers with similar methods [7, 54], was applied to analyze the usability performance data, while a one-way repeated measures ANOVA was used for sentiment ratings. Further details on this analysis will be presented in-line with the findings. For the qualitative analysis, we engaged with interview transcripts using an inductive approach based on grounded theory practices [15]. Borrowing from Charmaz's approach provided the flexibility to uncover emergent themes from the diverse experiences and reactions from our participants. Through a process of open coding and constant comparison, we systematically identified patterns and connections within and across our qualitative and quantitative data sets. This process allowed us to identify core themes related to user expectations, application context, and the perceived value or downsides of expressive design.

## 3 Findings

The results of our study provide compelling evidence that a return to fundamental design principles, as championed by expressive design, significantly improves mobile usability and user satisfaction. Our eye tracking results (RQ1) reveal that expressive designs lead to faster task completion and quicker user fixation on key elements, with a particularly pronounced benefit for older adults. This is complemented by our qualitative findings (RQ2), which uncover a nuanced understanding of why and when these design choices resonate with users. These insights demonstrate that expressive design is a powerful tool for creating interfaces that are more intuitive, efficient, and engaging across diverse application types.

### 3.1 Usability

To formally test our key hypothesis on the impact of expressive design on usability, we analyzed two key metrics collected during our experiment: 1) the time until the first eye fixation on the correct UI element as well as 2) the time to tap the correct UI element for each of the 20 screens presented to the participants.

*3.1.1 Fixation Time.* Descriptive statistics for the time to first fixation are shown in Figure 4A. On average, participants fixated on the key element more quickly in the expressive design conditions (M = 1.60 s, SD = 2.60 s) than in the baseline conditions (M = 2.40 s, SD = 4.53 s). This represents a 33% average improvement in eye fixation time for Expressive designs.

As illustrated in Figure 4B, this effect was most pronounced in the older age group. Participants aged 45 and older took more than twice as much time to fixate on the key element in the baseline conditions (M = 4.11 s, SD = 7.61 s) compared to the expressive conditions (M = 1.69 s, SD = 2.36 s). This difference, while still present, was substantially smaller for the younger age groups (18–24 years: M = 1.53 s vs. M = 1.31 s; 25–44 years: M = 1.97 s vs. M = 1.82 s, for baseline and expressive conditions, respectively).

To formally test the observed differences while accounting for individual-specific characteristics, we employed a fixed-effects linear regression model [2]. This model accounts for the repeated observations and controls for all time-invariant individual traits

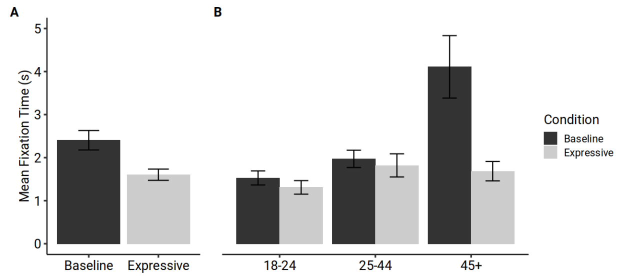


**Figure 4: Mean fixation time (in seconds) for Baseline and Expressive conditions. Panel (A) shows the overall main effect, and panel (B) shows the effect broken down by participant age group. Error bars represent the standard error of the mean.**

**Table 2: Fixed-Effects Regression Models Predicting Time to Fixation.**

| Predictor | Model 1 | Model 2 |
|---|---|---|
| **Direct Effects** | | |
| Condition [Expressive] | -.796$^{**}$ (.235) | -.219 (.174) |
| **Interaction Effects** | | |
| Expressive x Age Group [25-45] | | .081 (.288) |
| Expressive x Age Group [45+] | | -2.20$^{***}$ (.518) |
| **Model Statistics** | | |
| $R^2$ | 0.140 | 0.158 |
| N | 799 | 799 |
| Fixed Effects | Subject + App | Subject + App |

*Note:* Values are B (SE). The main effect of Age Group is absorbed by the participant fixed effects. The reference category for Condition is [Baseline] and for Age Group is [18-25]. +p < .10, $^{*}$p < .05, $^{**}$p < .01, $^{***}$p < .001.

(e.g., attention, digital literacy) through participant-level fixed effects and also accounts for any systematic difficulty associated with a specific app through app-level fixed effects. As a robustness check, we re-estimated the main models including screen order as a covariate to account for potential learning effects.

The results, presented in Table 2, confirm the descriptive findings. Model 1 shows a significant main effect of the design condition on time to first fixation (b = -.796, SE = .219, p < .01) indicating that the expressive design reduced fixation time by 0.796 seconds.

In Model 2, we introduced interaction terms between condition and age group. While the direct effect in the model turns out as not significant (b = -.219, SE = .174, p = .215), the analysis revealed a significant negative interaction for the 45-and-older age group (b = -2.20, SE = .518, p < .001), confirming that the beneficial effect of the expressive design was significantly larger for these older participants compared to the 18–24 reference group. In contrast, no significant interaction was found for the 25–44 age group (b = .081, SE = .288, p = .780), suggesting the age-related effect is specific to the 45+ participants. In other words, when controlling for age group, expressive design reduces the time to fixation for participants 45 and older by an average of 2.20 seconds, or 59%, matching the performance of younger age groups.

While the control for screen order revealed significant learning effects ($p < .001$), the main effects of interest remained statistically significant and unchanged in magnitude, demonstrating the robustness of our primary results.

When we look at specific applications, we see even stronger effects. For example, inviting a friend to a concert in Gmail saw a 73% reduction in median fixation time across age groups from 1.8 seconds to 0.48 seconds. Logging a manual activity in Fitbit saw a 71% reduction from 2.07 seconds to 0.6 seconds and snoozing an Alarm saw a 59% reduction from 0.98 seconds to 0.4 seconds.

*3.1.2 Tap Time.* On average, participants were also faster to tap the key elements in the expressive (M = 4.75s, SD = 8.74s) versions of the apps compared to the baseline ones (M = 5.92, SD = 8.35), an improvement of 20%. Descriptive statistics for the time to tap are visualized in Figure 5A.

Similar to our findings on fixation time above, participants 45+ saw the largest reductions in tap time in expressive designs (M = 6.51, SD = 12.78s) compared to baselines (M = 8.43s, SD = 11.46). The other age groups in our sample saw smaller reductions in tap time with the expressive designs over baselines (18–24 years: M = 3.65 s vs. M = 3.16 s; 25–44 years: M = 5.70 s vs. 4.60 s, for baseline and expressive conditions, respectively).

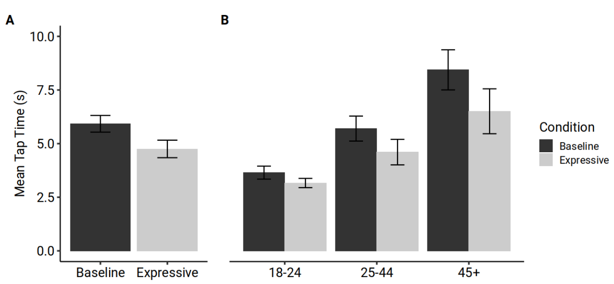


**Figure 5: Mean tap time to complete a task (in seconds) for Baseline and Expressive conditions. Panel (A) shows the overall main effect, and panel (B) shows the effect broken down by participant age group. Error bars represent the standard error of the mean.**

We used a set of fixed-effects linear regression models with the same specification as above to account for the data structure of the experiment results. We specified two models presented in Table 3 to test for the overall, direct effect of expressive design on tap time (Model 1) and the interaction of expressive design and age group to test the age effect (Model 2). Also in line with the above and as a robustness check, we re-estimated the main models including screen order as a covariate to account for potential learning effects.

For the direct effect of expressive design on tap time in Model 1, we find a statistically significant reduction in tap time (b = -1.17, SE = .410, p < .01). In other words, participants were able to tap the UI's correct element 1.17 seconds faster in the expressive designs versus the baseline designs. Looking at the interaction between condition and age group in Model 2, we find that both the direct

**Table 3: Fixed-Effects Regression Models Predicting Time to Tap**

| Predictor | Model 1 | Model 2 |
|---|---|---|
| **Direct Effects** | | |
| Condition [Expressive] | -1.17** (.410) | -.485 (.340) |
| **Interaction Effects** | | |
| Expressive x Age Group [25-45] | | -.595 (.700) |
| Expressive x Age Group [45+] | | -1.46 (1.04) |
| **Model Statistics** | | |
| $R^2$ | .233 | .235 |
| N | 917 | 917 |
| Fixed Effects | Subject + App | Subject + App |

*Note:* Values are B (SE). The main effect of Age Group is absorbed by the participant fixed effects. The reference category for Condition is [Baseline] and for Age Group is [18-25]. +p < .10, *p < .05, **p < .01, ***p < .001.

effect (b = -.485, SE = .340, p = .161) as well as the two interaction effects for 25-45 (b = -.595, SE = .700, p = .340) and 45+ (b = -1.45, SE = 1.04, p = .168) years old vs the reference group 18-24 to be non statistically significant.

While the control for screen order revealed significant learning effects ($p < .001$), the main effects of interest remained statistically significant and unchanged in magnitude, demonstrating the robustness of our primary results.

Again, looking at specific applications we found some with quite large effects. Inviting a friend to a concert in Gmail saw a 62% reduction in median tap time from 10.77 seconds to 4.05 seconds. Logging a manual activity in Fitbit saw a 25% reduction from 3.4 to 2.56 seconds and unmuting in Meet saw a 21% reduction from 3.0 to 2.37 seconds.

In summary, we found that across the 10 applications tested, Expressive designs were 33% faster for users to fixate on the correct UI element and 20% faster to tap on that element. These designs also equalized performance across ages, with users over 45 years of age improving their time to fixate by 59%, matching the performance of younger age groups. These effects are quite large, given the mature state of the baseline designs, and point to the possibility of large usability gains for a wide variety of applications.

## 3.2 User Sentiment

Our eye tracking results point to efficiency and learnability gains, two critical components of usability, brought about by Expressive design. However, there are additional factors to consider that are necessary to gauge the full user experience [41]. Specifically, we wanted to capture participants' satisfaction in response to the different designs. When participants were asked to rate each app design on a set of metrics, the Expressive screens were rated higher across all five aspects: Appealing, Clean, Fun, Modern, and Easiness. Figure 6 shows the results by condition for each of the sentiment metrics that participants rated during the experiment.

When comparing the means and standard deviations across these five metrics, differences emerge between conditions, most notably the 'Fun' and 'Modern' attributes. For instance, the mean rating for 'Fun' in the expressive condition (MExpressive = 3.97, SDExpressive

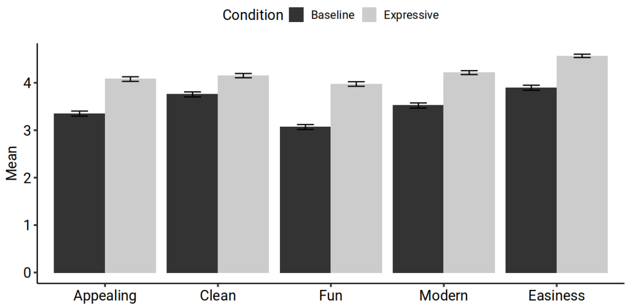


**Figure 6: Mean scores for each of the subjective metrics. In all, the Expressive designs scored more positively than the Baseline designs.**

**Table 4: Results of F-tests from Repeated Measure ANOVA for Each Metric.**

| Dependent Variable | df | F | p | $n^2p$ |
|---|---|---|---|---|
| Appealing | 1, 929 | 105.35 | < .001 | .102 |
| Clean | 1, 929 | 34.04 | < .001 | .035 |
| Easiness | 1, 929 | 118.19 | < .001 | .113 |
| Fun | 1, 929 | 169.46 | < .001 | .154 |
| Modern | 1, 929 | 105.15 | < .001 | .101 |

*Note:* The df values represent the between-groups and error degrees of freedom, respectively. The p-values are from one-way repeated measures ANOVAs comparing the Expressive and Baseline designs for each metric.

= 0.88) was substantially higher than in the baseline condition (MBaseline = 3.07, SDBaseline = 1.10). Similarly, a considerable gap was also observed for 'Modern' (MExpressive = 4.21 vs. MBaseline = 3.52) and 'Easiness' (MExpressive = 4.57 vs. MBaseline = 3.90). The smallest difference between the two conditions was for the 'Clean' metric (MExpressive = 4.15 vs. MBaseline = 3.76).

Validating these initial rankings we conducted a series of one-way repeated measures ANOVAs to compare the ratings of the Expressive and Baseline designs on these five attributes. This method accounts for the within-subjects design of our experiment and the resulting data structure. A repeated measures ANOVA is appropriate for this design as it accounts for the non-independence of observations and effectively partitions out the variance associated with individual participant differences. The results, summarized in Table 4, indicate that there was a statistically significant difference between the two designs for all attributes evaluated, substantiating our findings.

Looking at overall preference ratings for individual applications (see Figure 7), we ran a Binomial Test confirming that Expressive designs were significantly preferred with a threshold of p=0.05 in most of the apps, with two exceptions. One app, Docs, had very subtle design changes and there was no significant difference in which was preferred. The current version of the Setup app was the only app in which the Expressive design was not preferred, a point we will return to in the qualitative section below.

*3.2.1 Overall Preference for Expressive Designs.* Complementing the quantitative analysis above, in our interviews we found that users had a strong preference for the expressive designs. Many appreciated the colors and the simplicity that the changes in visual hierarchy brought to the screens. P05 told us, *"I love the colors, that's just how my personality and my soul is. Like I scream bold pink and beautiful and to the point, and it just seems very much made for me and it's very um it brings simplicity you know with regardless of what's going on. I just feel like it'll center me."* P41 expressed a similar sentiment: *"I'm more like an artistic person, creative, so I like it. I like the more color, the more fun, the more easier to navigate instead of this one's [Baseline] more dull, too much information. I like the color coding and the bigger fonts, more fun, even you know just looking at it, it's just more fun."*

Participants used words like *"expected," "simple,"* and *"bland"* to describe baseline designs which stood in contrast to vocabulary used to describe Expressive designs such as *"colorful," "exciting," and "modern."* When explaining their thoughts, most expressed that Expressive design dimensions such as color create a more fun experience. P26 explained their overall sentiments towards the two designs: *"Design B [baseline design] is definitely more boring. So if you want to see new things in your life, then starting fresh or colorful is like a good thing. It just makes you more happy."* Color was the most referred dimension in terms of its contribution to the experience. P08 said that *"those bright colors make it seem more fun...Our phone is so much a part of our lives now. So to have something at important to you in your life, it needs to have a little fun in it too."*

Several spoke specifically to usability improvements when commenting on all of the designs together. P03, who was in the 45+ age cohort, stated, *"I think it would potentially attract more older people with vision issues. It'd make them feel comfortable ... I like the color, I think that stands out the most, the pink, the blue. I think that is very helpful."* As seen above, those over 45 had the largest improvements in fixation time with the new designs. However, people of all ages talked about the *"easier, simple buttons"* (P46), and that the Expressive designs were *"a little more fun and clearer"* (P38). P23 spoke about the benefits for people with disabilities of a simpler visual hierarchy and larger controls: *"I'm also dyslexic so I'm going to pick whatever is visually the easiest for me."* Participants were excited for a potential launch of these designs. P1 stated *"that would be awesome…just mainly due to the visibility of the buttons on the screen… i like the color or the colors. The buttons [in the Expressive condition] were like lively, colorful."*

However some participants found aspects of the design that they did not prefer. Sometimes the bolder colors were too much for some participants. P38 stated, *"I think there are some aspects that feel a little bit over the top… where it just feels like there's a lot of colors and contrasts that aren't necessarily needed."* This was particularly true for the Setup screen, the only design that was preferred in the Baseline Material 3 version. P04 said that the Expressive Setup screen had *"too many colors... like foreground background and all that."* Because the Expressive design for Setup took up more space with instructions and a larger storage bar, some felt that the rest of the content was more *"scrunched in"* (P45). And P42 said that the Expressive version took *"a little bit too much space"* and preferred Design B because *"it feels like I can see more content in B even if that might not necessarily be the case."* Overall, participants strongly

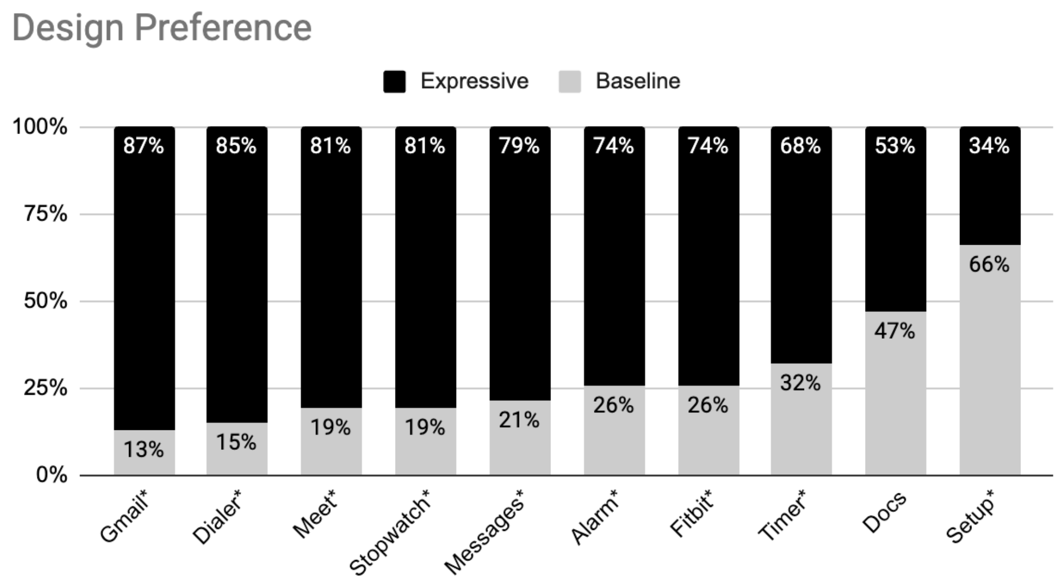


**Figure 7: Overall design preference by application. Eight apps saw significant preferences for the Expressive version, with one neutral, and one where the M3 Baseline was preferred (* indicates significance at the p=0.05 level or better).**

preferred the Expressive versions for the majority of apps, pointing to their increased usability, larger buttons, and bolder use of color.

*3.2.2 How Expressive Design Aided Usability.* Our qualitative analysis begins to unpack the nuances of when and where Expressive design features deliver maximum impact. It remains consistent that a user's perception and acceptance of usable designs are tied to the context and task [31, 50]. We found that the intentional use of design elements such as size, shape, color, and typography effectively guided user attention to areas of interest, impacting both perceived usability and aesthetic appeal. Our qualitative findings articulate how designers can successfully apply these expressive elements to support differing user goals—from simple, fast actions to complex, information-rich tasks.

*Color and size can draw focus when there are one or two key actions.* Often, a screen in an app will have one or two main actions that a user can perform, like snoozing or stopping an alarm or starting a stopwatch. When screens were simple and focused on completing an action quickly, participants appreciated bright colors and large buttons in enhancing visibility, supporting quicker task completion, which was reinforced by our eyetracking results. Representing the majority of participants, P14 reflected that *"the one that keeps standing out to me is the Stopwatch like that's something that I'm in the middle of a work out, I don't want to be stuck looking at my phone the whole time. I want to act quick."*

The need for accuracy and fast response times was enabled by the larger button sizes for less complex screens enabled by Expressive design. P40 emphasizes this point by sharing that they liked *"the start button, how it's big and easier to press because of how big it is... if it's smaller it can be hard to stop and start sometimes because you can click next to or above the button."* These design differences can be clearly seen in Figure 8. Similar to the Stopwatch, the use of a bold color and large button size was also preferred by our participants for snoozing an alarm in the Alarm app. P45 stated, *"You have a giant button so you just you know you could like close your eyes and hit it even."* In both scenarios participants discussed that being able to act quickly reduced the potential for human error by directing focus and attention to a singular task.

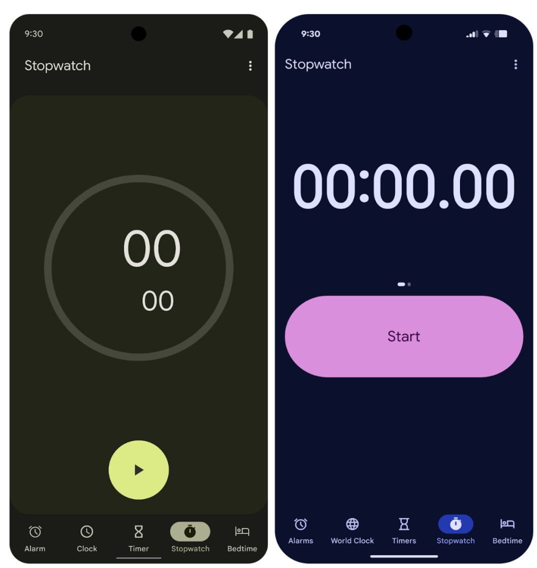


**Figure 8: The design of the stopwatch app put a much larger emphasis on the Start button by using size, shape, and a text label to highlight this action.**

However, creating more usable experiences with expressive design also means contending with the responsibilities of building and maintaining user trust. When describing the potential benefits of larger button sizing and color, participants began to recognize the potential impacts of design on their agency and decision making. We can see this tension emerge in P47's response to the increased size of the snooze alarm button where they shared *"I feel like it wouldn't take me an extra second to press snooze but I feel like if someone would have that, they would be more likely to hit snooze because it's like the biggest button on the screen."* An important factor of Expressive design is that size can drive the user's attention towards an item they might not want, a topic we will engage more deeply in the Discussion.

In more complex app designs, such as Timer and Fitbit, we found that color worked well to highlight an action in what participants referred to as *"messy"* backgrounds due to the number of timers and fitness milestones visually represented. Through the use of a pop of contrasting color in the expressive design, participants were able to locate the "+" button" because the color used stood out from the rest of the screen. P23 described this experience: *"If I'm looking to add a new timer, there is a lot of content to scroll through. The redeeming factor of E [expressive design] was that it's like adding a new timer was a different coloration to it."* While P04 similarly stated, *"Well design E [the expressive version], I like the fact that it's [+ button] larger. I also like how they are using a pop color, that the only other place where I see it is on the main center icon."* In this circumstance, directing attention and focus with Expressive design cues like size and color were not seen as distracting, but as essential tools helping focus attention ultimately supporting faster task completion.

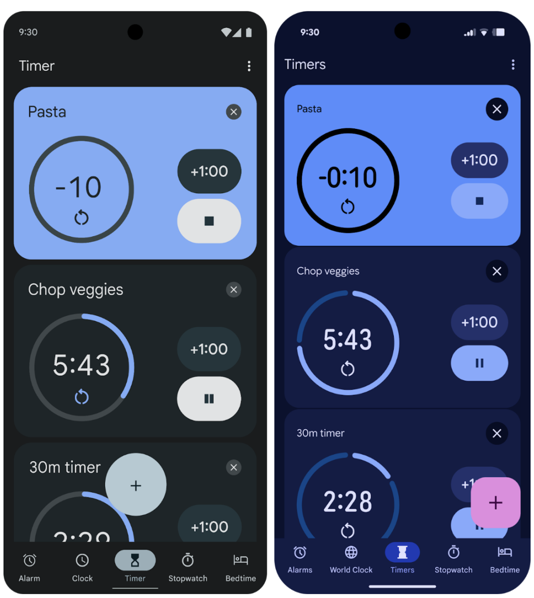


**Figure 9: The design of the Timer app, which used a contrasting color for the "+" Floating Action Button to create a new timer.**

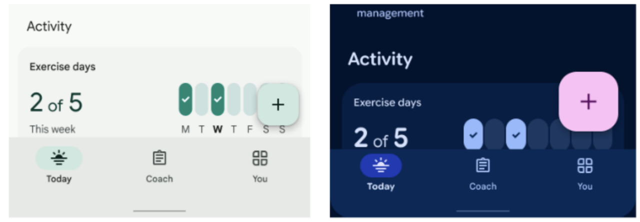


**Figure 10: The change in size and color for the "Log New Activity" button (+) in Fitbit.**

In information-rich apps like Meet and Messages, color takes on a supportive role in guiding users to secondary functions, buttons used less frequently. The priority for these apps is screen space optimization for two different reasons: 1) to focus on the content (e.g., the face of the other person or reading messages) and 2) having tools that could be used to complete the overall experience readily available without being distracting. These apps require more intentional use of color in order to have tools readily available to the user to complete their overall task. In the Messages app, participants were able to locate the audio message button quickly because they *"can just go straight to the green button and I know that is what I am looking for"* (P09). While not being the main focus or the most common task of typing a message, color was able to call attention to a secondary function without being distracting.

Overall, the use of color and size can support visibility around primary and secondary-focused buttons. We saw this in apps such as Alarm and Stopwatch where the use of bold colors and large buttons enhanced the experience. By highlighting primary-focused buttons in busier interfaces, the use of contrasting colors and large buttons supported faster identification and completion of the "Log New Activity" task in the Fitbit app. The importance of variable color and size was reinforced in information-rich apps like Messages, where secondary buttons such as the "Audio Message" were able to remain easily discoverable without being overbearing.

*Containment and typography can enhance usability in more complex screens.* In addition to using size and color to manage complexity in screens with a large number of actions, our study illustrates the value of containment and typography as Expressive design techniques. Color-based containment reinforces information hierarchy and guides the users' focus by organizing the screen content into logical groupings. The particular use of containment in the Google Meet app was appreciated by participants because it separated the controls from the video of the call as well as moved the hang up button farther away from the mute button, two actions that are commonly confused when quickly trying to come off of mute to speak. P10 mentioned that they, *"like how the buttons aren't just free-standing, they are segmented on the black bar. Feels more organized."* P13 told us that *"floating icons can be difficult to spot so having the icons in a bar helps."* In this example, the size and tint of the larger mute button was also able to help guide the user's attention to the most common action in the action bar.

In Gmail and Dialer, containment and typography supported participants in scanning their screens to locate the body of information needed. P31 *"like[d] the fact that it's really quick to pick out the date"* in the Gmail Expressive version. *"The information is kind of set better because there is a lot of information."* Similarly, in the

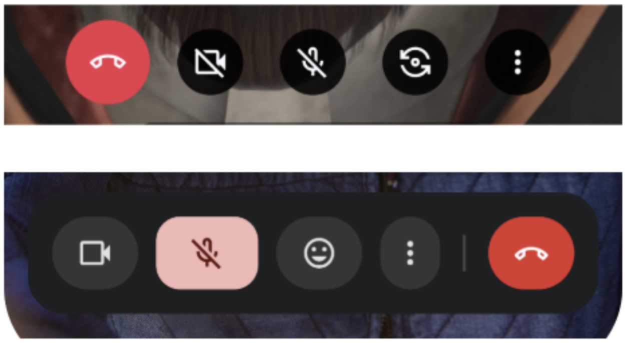

**Figure 11: A comparison of the action bars in the Baseline (top) and Expressive (bottom) versions of Meet. In the Expressive version, the Mute control uses color, size, and shape to distinguish itself from the other controls while providing physical separation from the hang up button to avoid accidental taps.**

Dialer Expressive design, the bar to answer or decline a call *"pops out, I like how I can see the decline and answer. It's not as small as the one on B. Everything is more in plain view for me"* (P09). In situations with many actions, containment, color, and typography allow for scannability to easily locate necessary information and functions.

Overall, participants noticed the visual changes in the applications and the use of color, size, and containment to guide their attention when completing a task. For screens with many actions, containing them in a single section and using color to highlight important actions helped users to quickly make sense of many options. For more simple screens, large and colorful controls quickly guided users to successfully accomplishing the main action. The variations enabled by Expressive Design emphasize the potential for continued usability gains while underscoring the central role of designers to use the appropriate tactics to guide user attention.

## 4 Limitations

There are a few limitations of this study. First, the tasks that users were given all involved only a single tap to complete. Real world tasks in mobile applications often involve several screens and interactions with a variety of components to complete. However we chose single tap tasks as these are best suited for eyetracking analysis and the fixation time questions that we had as well as the instrument design of randomizing the screens. As mentioned above, we chose a variety of tasks that were the primary action on the screen (e.g. starting a timer, answering a call) and some that were secondary actions (e.g. bolding text, un-muting a video call) to best cover a variety of use cases. Future research should explore more complex workflows to understand how expressive principles impact usability over a complex multi-screen task.

Also, this study only focused on mobile interfaces as displayed on a single device. We did not test screens on desktops, tablets, foldables, XR, or other environments. With this work we cannot say whether Expressive design tactics will work as well on these other devices which provides a clear opportunity for additional research and next steps.

Our study was conducted in a lab setting with the device positioned at a fixed distance from the participant. This is the optimal setup for eye tracking as real world conditions can interfere with the calibration of the device. While we have no evidence to think that the lab setting would affect differences in time to fixate or tap, real world evaluations in real contexts of use (on the subway, walking down the street, etc.) could provide additional situational validation of increased usability.

The study was also conducted in only one city. While a diverse set of ages, genders, and ethnic backgrounds were recruited, cultural differences in other parts of the world could impact aesthetic preferences or performance during interactions. While 48 participants is a fairly large sample for an in-lab study using eye tracking, an even larger sample could have given us increased power to find additional effects.

## 5 Discussion

Through this study we have seen that there is a large opportunity for even some of the most researched and usable apps to become substantially more usable. Across the 10 designs tested, we saw significant decreases in both time to fixate on the correct component for a task and overall task completion time. Participants were 33% faster to spot and 20% faster to tap the correct element on the screen across designs. The added flexibility of the design system to offer larger component sizes, contrasting colors, containment, and variable typography provided a significantly more usable interface.

In addition, participants strongly preferred the new designs overall and on five different metrics of design aesthetics. The designs were not only more usable, but they provided a new visual language that was preferred over the rigid defaults of Material 3 Baseline. When every button is the same size and every screen organized in the same way, the experience of using an app can be quite boring. These key findings have a variety of implications for design system creators and app designers.

### 5.1 The march towards better usability continues

While Nielsen argues that additional usability gains will be hard fought, and perhaps not worth the return on investment for most teams [42], we have found significant usability gains can be achieved by returning to the basic principles of design as enabled by the latest design systems. Using size, shape, color, containment, and typography can make key actions stand out and can reduce eye scanning time in an interface by grouping similar actions or content together.

Returning to two of the fundamental laws of HCI, we have found ways to increase usability while still providing the same number of options on the screen. The Hick-Hyman Law [14] states that more options on the screen will lead to slower interaction times. However, not all options need to be shown with equal saliency. We've seen that applying color, size, shape, or containment to primary and secondary actions can reduce interaction time while keeping the number of options on the screen fixed. Users will fixate on these elements more quickly and not need to scan every item on the screen for common actions. Similarly Fitts' Law [23] states that larger UI components will lead to faster interaction times. This continues to be true with our data.

Design has been caught in a local maximum, working within very prescriptive design systems that set fixed sizes for components and used minimal color differences to add emphasis. Systems such as Material 3 Expressive show that a much wider range of flexibility can lead to increased usability and preference for designs.

Designers can learn from this study and use these tactics to better guide user attention in their applications whether they use Material Design or not. Given that we studied 10 very diverse applications, from email to fitness tracking to alarm clocks, we have confidence that these principles could be applied to a wide array of applications to improve their usability and overall desirability. This can be a step change in overall system usability, and not just the small incremental gains that many believe are all that is possible given the state of usability today.

## 5.2 Design Systems should offer additional flexibility

Over the years, Design Systems have favored uniformity over the feeling a user has while using an application. While resulting apps that strictly follow the Apple HIG or Google Material 3 are usable and familiar, they tend to look and feel the same, and as we have seen in this study are not as usable as they could be with additional flexibility.

This study has shown that there can be benefits to offering wider ranges of options for designers in a design system. When every button is Material Blue and 48 dp tall, actions feel the same level of priority and users must scan across the screen to find what they are looking for. But for screens that have a true primary action, increasing the size of these buttons or altering their color can make them much easier to spot and faster to tap. This holds for secondary actions as well, where a little contrasting color or extra containment can add visual saliency.

This variation also provides a wider visual range across the applications that a user interacts with, which can bring greater feeling into each application to match its purpose. A photo browsing app doesn't need to look and feel like a corporate calendar app. Yet they can follow from the same broader design system, where overall patterns can be learned in one app and still used in others. While providing additional size or shape variants for components, a single stylistic design can still apply across the system for the consistency that is important to most design systems. Design system creators should not be afraid of variety in size, color, or shape, as long as the system overall provides coherent components and patterns.

## 5.3 The Responsibility of Design

The tools of Expressive design enable greater customization and autonomy for designers to create more engaging user experiences. Our quantitative results clearly demonstrate that this approach improves user focus and attention, while our qualitative inquiry provides a nuanced understanding of how. Yet, we need to critically engage with the ethical responsibilities that accompany this increased flexibility. Design is inherently persuasive [40, 43, 49]. As Gray et al. describe, "persuasive tech is often praised for the good it is capable of producing in society and individual life such as encouraging socially responsible behavior or the bettering of personal habits; there are also substantial ethical considerations" [29]. Embracing this perspective shapes our commitment to promoting ethical UX design by pointing out the tensions and potential for misuse.

The very elements that our study found to improve usability, such as size, color, containment, and typography, are also the foundational elements that enable what has been termed deceptive patterns. Originally coined by Brignull [13] as "dark patterns" to describe dubious design approaches intended to trick users, the term has more recently evolved to "deceptive patterns" to broaden the types of patterns that can mislead users and to address the racial undertones of the original term. These manipulative typologies—including nagging, obstruction, sneaking, interface interference, and forced action [29]—are enabled by the same design choices that our study demonstrates can lead to a more usable interface.

Our findings, which show that larger buttons and contrasting colors reduce fixation and tap times, illustrate how these small, micro-decisions can powerfully influence user behavior. For example, while a large, distinct "Start" button on a stopwatch helps users complete their task faster, a designer could just as easily use a large, colored button to encourage a task that is not in the user's best interest. We saw this tension reflected in participant comments, such as P47's recognition that the large Alarm snooze button, while highly usable, also subtly persuades the user to hit snooze.

While deceptive design patterns historically have been geared towards gaining access to data or increased monetary spending and time spent on mobile apps, video games, or online shopping, the persuasive nature of these techniques is widespread [18, 38]. Manipulation has been proven to be a problem for users of all ages, from adults [33] to pre-teens [51] and children [48]. Recent work also shows a connection to social safety, highlighting that deceptive patterns are not solely associated with financial outcomes [16] but can result in much broader societal impacts.The double edge sword of design is that while these tactics can be used maliciously, the same tools of color, shape, size, and placement can also undo deceptive patterns in tactics called Bright Patterns [12].

We argue that the increased flexibility of expressive design deepens a designer's responsibility to consider the ethical implications of their choices but also offers more ways to counter deceptive designs. Through our work we aim to raise awareness of the critical role that seemingly simple design decisions play in enabling both highly usable interfaces and potentially deceptive patterns. Research has shown that users can identify deceptive design patterns when appropriately informed [18], the same should apply to designers.

In addition, while proscriptive design systems may have placed a 'ceiling' on usability, they also provided a 'floor' for the vast majority of apps and teams, ensuring that most apps would meet a baseline level of usability. The shift to Expressive design systems must also be accompanied by an increase in the quality of education and guidelines to avoid 'backsliding' in usability or deceptive tactics for interfaces that are created by novice designers or engineers without design training, as many commercial apps are today.

# 6 Conclusion

We have conducted a first of its kind analysis of designs produced using the new Material 3 Expressive design system. Through examining ten applications designed with Expressive and Baseline Material 3 versions, we found that Expressive designs were 33% faster for users to fixate on the correct UI element for a task and 20% faster to tap on that element, with users over 45 years of age improving their time to fixate by 59%, matching the performance of other age groups. Users also significantly preferred eight of the ten Expressive designs, and rated Expressive designs significantly higher on a set of five aesthetic attributes.

This study has demonstrated that there is quite some room for mobile usability to improve, in contrast with recent opinions of usability professionals. The constraints of existing design systems were holding designs in a local maximum. However, returning to design basics of using size, shape, color, and containment to help guide user attention can create interfaces that are more usable and preferred by a diverse sample of users. Design systems can still provide consistency, but through a larger number of stylistically aligned variants per component.

We hope that this study encourages mobile designers as well as those creating design systems to lean towards more flexibility in their designs to create more optimized designs for specific types of interactions, designs which contain more feeling and can help users easily find the controls needed to complete their tasks. Expressive design tactics are not just about aesthetics but can provide significantly more usable interfaces for all.

# Acknowledgments

We would like to thank our Google product designers for contributing the screens that were tested as well as the entire Material Design team for their work to define Material 3 Expressive. Thanks also to Matías Duarte and Vanessa Cho for early feedback on the paper and framing in the history of design systems.

# References

[1] National Aeronautics and Space Administration. 1976. *National Aeronautics and Space Administration Graphics Standards Manual.* NASA.
[2] Paul D Allison. 2009. *Fixed effects regression models.* SAGE publications.
[3] Apple. 2025. Buttons. https://developer.apple.com/design/human-interface-guidelines/buttons.
[4] Apple. 2025. Human Interface Guidelines. https://developer.apple.com/design/human-interface-guidelines.
[5] Inc Apple Computer. 1987. *Apple human interface guidelines: The Apple desktop interface.* Addison Wesley Publishing Company.
[6] Aiman M. Ayyal Awwad, Christian Schindler, Kirshan Kumar Luhana, Zulfiqar Ali, and Bernadette Spieler. 2017. Improving pocket paint usability via material design compliance and internationalization & localization support on application level. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 99, 8 pages. doi:10.1145/3098279.3122142
[7] Melanie Bancilhon, Alvitta Ottley, and Andrew Jordan. 2025. The Anatomy of a Plea: How Uncertainty, Visualizations & Individual Differences Shape Plea Bargain Decisions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1177, 18 pages. doi:10.1145/3706598.3713096
[8] Jennifer Romano Bergstrom and Andrew Schall. 2014. *Eye tracking in user experience design.* Elsevier.
[9] Susanne Bødker. 2006. When second wave HCI meets third wave challenges. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles* (Oslo, Norway) *(NordiCHI '06)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/1182475.1182476
[10] Aga Bojko. 2013. *Eye tracking the user experience: A practical guide to research.* Rosenfeld Media.
[11] Letizia Bollini et al. 2016. From skeuomorphism to material design and back. The language of colours in the 2nd generation of mobile interface design. *Colour and colorimetry. Multidisciplinary contributions* 12 (2016), 309–320.
[12] 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 *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 763–776. doi:10.1145/3461778.3462086
[13] Harry Brignull. 2025. Deceptive Patterns. https://www.deceptive.design/.
[14] Stuart K. Card, Allen Newell, and Thomas P. Moran. 1983. *The Psychology of Human-Computer Interaction.* L. Erlbaum Associates Inc., USA.
[15] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis.* sage.
[16] Ishita Chordia, Lena-Phuong Tran, Tala June Tayebi, Emily Parrish, Sheena Erete, Jason Yip, and Alexis Hiniker. 2023. Deceptive Design Patterns in Safety Technologies: A Case Study of the Citizen App. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 193, 18 pages. doi:10.1145/3544548.3581258
[17] Santiago de Leon-Martinez, Robert Moro, Branislav Kveton, and Maria Bielikova. 2025. Riding the Carousel: The First Extensive Eye Tracking Analysis of Browsing Behavior in Carousel Recommenders. *arXiv preprint arXiv:2507.10135* (2025).
[18] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. 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. doi:10.1145/3313831.3376600
[19] Marcin Drygielski, Agnieszka Indyka-Piasecka, Mateusz Piwowarczyk, Zbigniew Telec, Bogdan Trawiński, and Trong Hai Duong. 2019. Usability Testing of Data Entry Patterns Implemented According to Material Design Guidelines for the Web. In *Computational Collective Intelligence*, Ngoc Thanh Nguyen, Richard Chbeir, Ernesto Exposito, Philippe Aniorté, and Bogdan Trawiński (Eds.). Springer International Publishing, Cham, 697–711.
[20] Andrew T Duchowski and Andrew T Duchowski. 2017. *Eye tracking methodology: Theory and practice.* Springer.
[21] Umer Farooq and Joseph T. Munko. 2015. Industry Is Changing, and So Must We. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 655–661. doi:10.1145/2702613.2702962
[22] Desika Anggi Fitriana, Irwan Iftadi, Rahmaniyah Dwi Astuti, et al. 2023. User experience evaluation and interface redesign using user experience questionnaire and Google material design (case study: SIAKAD UNS). *Jurnal Teknik Industri* 13, 2 (2023), 159–164.
[23] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.
[24] Adrian Forty. 1979. Lorenzo of the Underground. *The London Journal* 5, 1 (1979), 113–119. arXiv:https://doi.org/10.1179/ldn.1979.5.1.113 doi:10.1179/ldn.1979.5.1.113
[25] Stephen Giff and Huseyin Dogan. 2016. Ux research is dead. Long live ux research!. In *Proceedings of the 30th International BCS Human Computer Interaction Conference.* BCS Learning & Development.
[26] Google. 2014. Material Design. https://m1.material.io.
[27] Google. 2025. Buttons, Specs. https://m3.material.io/components/buttons/specs.
[28] Google. 2025. Material Design 3. https://material.io.
[29] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3173574.3174108
[30] Robert J.K. Jacob and Keith S. Karn. 2003. Commentary on Section 4 - Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The Mind's Eye*, J. Hyönä, R. Radach, and H. Deubel (Eds.). North-Holland, Amsterdam, 573–605. doi:10.1016/B978-044451020-4/50031-1
[31] Jeff Johnson. 2020. *Designing with the mind in mind: simple guide to understanding user interface design guidelines.* Morgan Kaufmann.
[32] Amy K Karlson, Benjamin B Bederson, and J Contreras-Vidal. 2006. Understanding single-handed mobile device interaction. *Handbook of research on user interface design and evaluation for mobile technology* 1 (2006), 86–101.
[33] Woon Chee Koh and Yuan Zhi Seah. 2023. Unintended consumption: The effects of four e-commerce dark patterns. *Cleaner and Responsible Consumption* 11 (2023), 100145. doi:10.1016/j.clrc.2023.100145
[34] Hsiao-Chun Lai and Rushani Wirasinghe. 2017. Applied Research for Advertising Products: Tactics for Effective Research. In *Proceedings of the 2017 CHI Conference*

*Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 1144–1151. doi:10.1145/3027063.3053338

[35] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold P.O.S. Vermeeren, and Joke Kort. 2009. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 719–728. doi:10.1145/1518701.1518813

[36] Craig M. MacDonald, July Sosebee, and Alexandra Srp. 2022. A Framework for Assessing Organizational User Experience (UX) Capacity. *International Journal of Human–Computer Interaction* 38, 11 (2022), 1064–1080. arXiv:https://doi.org/10.1080/10447318.2021.1979811 doi:10.1080/10447318.2021.1979811

[37] Nurul Hidayah Mat Zain, Fariza Hanis Abdul Razak, Azizah Jaafar, and Mohd Firdaus Zulkipli. 2011. Eye Tracking in Educational Games Environment: Evaluating User Interface Design through Eye Tracking Patterns. In *Visual Informatics: Sustaining Research and Innovations*, Halimah Badioze Zaman, Peter Robinson, Maria Petrou, Patrick Olivier, Timothy K. Shih, Sergio Velastin, and Ingela Nyström (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 64–73.

[38] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 360, 18 pages. doi:10.1145/3411764.3445610

[39] Rolf Molich and Jakob Nielsen. 1990. Improving a human-computer dialogue. *Commun. ACM* 33, 3 (March 1990), 338–348. doi:10.1145/77481.77486

[40] Harold G Nelson and Erik Stolterman. 2014. *The design way: Intentional change in an unpredictable world*. MIT press.

[41] Jakob Nielsen. 2012. Usability 101: Introduction to Usability. https://www.nngroup.com/articles/usability-101-introduction-to-usability/.

[42] Jakob Nielsen. 2025. Declining ROI From UX Design Work. https://jakobnielsenphd.substack.com/p/low-ux-roi.

[43] Chris Nodder. 2013. *Evil by design: Interaction design to lead us into temptation*. John Wiley & Sons.

[44] Jean Carlos Alves Duarte Passos. 2024. Estudo sobre a aplicação de UX design no desenvolvimento de interfaces interativas usando material design. (2024).

[45] Sebastian A. C. Perrig, David Ueffing, Klaus Opwis, and Florian Brühlmann. 2023. Smartphone app aesthetics influence users' experience and performance. *Frontiers in Psychology* Volume 14 - 2023 (2023). doi:10.3389/fpsyg.2023.1113842

[46] Aryo Pinandito, Hanifah Muslimah Az-zahra, Lutfi Fanani, and Anggi Valeria Putri. 2017. Analysis of web content delivery effectiveness and efficiency in responsive web design using material design guidelines and User Centered Design. In *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*. 435–441. doi:10.1109/SIET.2017.8304178

[47] Alex Poole and Linden J Ball. 2006. Eye tracking in HCI and usability research. In *Encyclopedia of human computer interaction*. IGI Global Scientific Publishing, 211–219.

[48] Jenny Radesky, Alexis Hiniker, Caroline McLaren, Eliz Akgun, Alexandria Schaller, Heidi M. Weeks, Scott Campbell, and Ashley N. Gearhardt. 2022. Prevalence and Characteristics of Manipulative Design in Mobile Applications Used by Children. *JAMA Network Open* 5, 6 (06 2022), e2217641–e2217641. doi:10.1001/jamanetworkopen.2022.17641

[49] Johan Redström. 2006. Persuasive Design: Fringes and Foundations. In *Persuasive Technology*, Wijnand A. IJsselsteijn, Yvonne A. W. de Kort, Cees Midden, Berry Eggen, and Elise van den Hoven (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 112–122.

[50] Katharina Reinecke and Abraham Bernstein. 2011. Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. *ACM Trans. Comput.-Hum. Interact.* 18, 2, Article 8 (July 2011), 29 pages. doi:10.1145/1970378.1970382

[51] René Schäfer, Sarah Sahabi, Annabell Brocker, and Jan Borchers. 2024. Growing Up With Dark Patterns: How Children Perceive Malicious User Interface Designs. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction* (Uppsala, Sweden) *(NordiCHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 25, 17 pages. doi:10.1145/3679318.3685358

[52] Ben Shneiderman. 2010. *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India.

[53] G. Lynn Shostack. 1984. Designing services that deliver. *Harvard Business Review* 62, 1 (1984), 133–139.

[54] Sofia Eleni Spatharioti, David Rothschild, Daniel G Goldstein, and Jake M Hofman. 2025. Effects of LLM-based Search on Decision Making: Speed, Accuracy, and Overreliance. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1025, 15 pages. doi:10.1145/3706598.3714082

[55] statcounter GlobalStates. 2025. Mobile Operating System Market Share Worldwide. https://gs.statcounter.com/os-market-share/mobile/worldwide.

[56] Maria Stone, Frank Bentley, Brooke White, and Mike Shebanek. 2016. Embedding User Understanding in the Corporate Culture: UX Research and Accessibility at Yahoo. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 823–832. doi:10.1145/2851581.2851592

[57] Amalia Suzianti and Anggar Belahakki. 2020. Redesigning User Interface of MRT Jakarta's Mobile Application using Usability Testing Approach. In *Proceedings of the 6th International Conference on Industrial and Business Engineering* (Macau, Macao) *(ICIBE '20)*. Association for Computing Machinery, New York, NY, USA, 73–78. doi:10.1145/3429551.3429587

[58] Tobii. 2025. Tobii Pro Glasses 3. https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3/scientific-use-cases.

[59] Linn Vizard. 2015. The Evolution of Buttons in UX Design. https://blog.adobe.com/en/publish/2015/10/19/the-evolution-of-buttons-in-ux-design.

[60] Jiahui Wang, Pavlo Antonenko, Mehmet Celepkolu, Yerika Jimenez, Ethan Fieldman, and Ashley Fieldman. 2019. Exploring Relationships Between Eye Tracking and Traditional Usability Testing Data. *International Journal of Human–Computer Interaction* 35, 6 (2019), 483–494. arXiv:https://doi.org/10.1080/10447318.2018.1464776 doi:10.1080/10447318.2018.1464776

[61] Meltwater we are social. 2024. Digital 2024 Global Overview Report. https://datareportal.com/reports/digital-2024-global-overview-report.

[62] Ruo-Qiao Zhao and Li-Hao Chen. 2020. Research on Interface Design for the Elderly. In *HCI International 2020 - Posters*, Constantine Stephanidis and Margherita Antona (Eds.). Springer International Publishing, Cham, 128–135.

[63] Jakub Štěpán Novák, Jan Masner, Petr Benda, Pavel Šimek, and Vojtěch Merunka. 2024. Eye Tracking, Usability, and User Experience: A Systematic Review. *International Journal of Human–Computer Interaction* 40, 17 (2024), 4484–4500. arXiv:https://doi.org/10.1080/10447318.2023.2221600 doi:10.1080/10447318.2023.2221600