Unconscious Frustration: Dynamically Assessing User Experience using Eye and Mouse Tracking

SCOTT A. STONE*, Department of Psychology, Neuroscience and Mental Health Institute, University of Alberta, Canada
CRAIG S. CHAPMAN†, Faculty of Kinesiology, Sport and Recreation, Neuroscience and Mental Health Institute, University of Alberta, Canada

Eye-tracking has become easier to deploy in user experience (UX) studies to get a sense of where users attend to during interactions. Additionally, mouse tracking grants insights into the cognition driving the user's behaviours and end goals, as can measuring the coordination between the eye and mouse-cursor. We created a menu navigation task based on a popular video game to assess two populations: a local cohort, and a remote cohort. We used two different eye trackers (monitor-mounted hardware, and a webcam-based algorithm; local used both simultaneously, remote used webcam only) with concurrent mouse tracking to detect friction in the UX. We found that both eye trackers had similar performance and revealed a previously undetected friction point. We argue this friction point was only detected because of the use of quantified, coordinated unconscious behaviours (eye and hand movements). The methods demonstrated are easily integrated into current UX studies with minimal cost.

CCS Concepts: • **Human-centered computing** → **Systems and tools for interaction design**; **Human computer interaction (HCI)**; **Usability testing**.

Additional Key Words and Phrases: eye tracking, mouse tracking, user experience

ACM Reference Format:
Scott A. Stone and Craig S. Chapman. 2023. Unconscious Frustration: Dynamically Assessing User Experience using Eye and Mouse Tracking. *Proc. ACM Hum.-Comput. Interact.* 7, ETRA, Article 168 (May 2023), 17 pages. https://doi.org/10.1145/3591137

INTRODUCTION

Background: user experience

Measuring user experience (UX) is difficult and many of the methods employed in UX research (UXR) are qualitative [Hinderks et al. 2019; Law et al. 2014]. These approaches have obvious value, but make the critical assumption that the end-user is consciously aware of their own experiences. However, users can have a poor experience and not have cognitive access to the reasons why. For example, a user could be unconsciously focused on an irrelevant but visually salient part of a menu but not be able to recognize that distracted behaviour upon later reflection. As a result, users can be confused for reasons they may not be able to articulate. Using traditional qualitative methods, this unconscious frustration is hard (or impossible) to measure. While quantitative research has

Authors' addresses: Scott A. Stone, sastone@ualberta.ca, Department of Psychology, and Neuroscience and Mental Health Institute, and University of Alberta, Edmonton, Alberta, Canada; Craig S. Chapman, c.s.chapman@ualberta.ca, Faculty of Kinesiology, Sport and Recreation, and Neuroscience and Mental Health Institute, and University of Alberta, Edmonton, Alberta, Canada.

2573-0142/2023/5-ART168 $15.00
https://doi.org/10.1145/3591137

been conducted in UXR, few solid methodologies exist in the field [Hauger et al. 2011; Huang et al. 2012; Papoutsaki 2015; Papoutsaki et al. 2017]. Here, we investigate if a systematic quantitative approach to measuring unconscious behaviour by simultaneously recording eye gaze and mouse movements improves our interpretation of user behaviour.

Background: tracking the eyes and hands

As highlighted in the example above, where you look can be tightly linked to your experience but often falls below a threshold of awareness. Therefore, tracking a person's gaze is an important tool for measuring a person's unconscious experience. Fortunately, eye tracking has become more portable and affordable in recent years. An eye tracker records the position of the pupils, which can be transformed into a projected gaze position. Eye movements can inform us about the cognitive processes used to interact with our environment [Hayhoe 2000; Hayhoe and Ballard 2005; Land and Hayhoe 2001]. For example, Land and Hayhoe analyzed eye movements while participants completed a simple tea- or sandwich-making task. In general, the eye movements encoded the necessary steps to complete the task, and a generalized model was capable of predicting the order of behaviours. The increased portability of newer eye trackers has made it easier to extend eye-tracking studies outside of the laboratory. For example, monitor-mounted gaze tracking devices can be had for as little as two hundred US dollars and the advent of webcam eye tracking has made it possible to collect data from almost anyone in the world right from their own home. Due to this widespread availability, webcam eye tracking is specifically useful in the field of UXR as it grants access to the context UX researchers are most interested: naturalistic interaction. However, what is not clear is whether a webcam eye tracker can be used as a drop-in replacement for a more conventional monitor-mounted eye tracker given that it is likely to have lower temporal and spatial resolution.

While eye tracking can provide information about what a person may be seeking or distracted by, eyes typically are not used as an input device. Therefore, to understand the chain from information to input that ultimately results in a particular experience, we also need to measure how people interact with their workspace. For digital UX, the most common input device is a computer mouse, whose movements can be easily collected during interaction. Measuring hand movements grants insight into the dynamics of cognition (for a review, see [Freeman et al. 2011]). Freeman et al. argue that motor and cognitive systems in the brain are not independent, but rather deeply intermingled. Movements are continuously updated through visual cognition over time [Goodale 2011; Goodale and Milner 1992; Goodale et al. 1986; Milner and Goodale 2006]. Recording even simple hand movements can therefore offer deep insight into cognitive processes and how they dynamically unfold [Chapman et al. 2010; Gallivan and Chapman 2014; Gallivan et al. 2018; Stone et al. 2022; Wispinski et al. 2018]. Simultaneous eye and mouse tracking may therefore grant deeper insight into the UX assessment process than previous methods.

Related works: eye and hand tracking in UXR

A primary goal of UXR is to understand what produces high quality experiences that promote end-user adoption (for a detailed review, see [Nunnally and Farkas 2016]). A major barrier to adoption is when a user experiences a friction point - an interaction that confuses or slows down the user. A useful test bed for friction points, and therefore a good candidate for us to test the collection of eye and hand movement data, is a user interface (UI). UIs are one of the primary ways that we interact with computer software, and how users navigate through UIs presents many opportunities for both conscious and unconscious friction points to arise. Because the eyes and the hands are the primary medium for users to interact with the UI, recordings will encode points of friction, but it can be tricky to interpret and analyze the data. One method is to collapse the data across time to get a static look at user performance. For example, when interacting with web pages, users tend to look at the areas they are about to interact with [Guo et al. 2013; Hauger et al. 2011; Huang et al. 2012; Papoutsaki 2015; Papoutsaki et al. 2017]. An intuitive way to visualize static data is known as a heatmap, which requires collapsing across time to calculate averaged statistics. But, static data interpretations like heatmaps fail to capture real human behaviours which are typically dynamic, with reach and gaze trajectories being updated in real-time [Goodale 2011; Westwood and Goodale 2003].

Dynamic measures, then, may be best suited to accurately detect friction points, especially those of which the user is not aware. One way to collect coarse dynamic measures is by splitting the data into natural phases. Here, a phase can coincide with the user's current goal. For example, if a user was attempting to send an email on a web page, the task could ostensibly be split into three phases: 1) locating the button to compose an email, 2) composing the email and 3) locating and clicking on the send button. Friction can be detected in any or all of the phases, but if an analysis collapses across the entire task the researcher will be puzzled as to *when* the actual friction occurred. Navigating through a UI is similar and their structure therefore offers a useful scaffolding onto which we can structure an analysis of behaviour. A user will have a goal of the menu object they are trying to get to, and getting there will require the user to complete several 'sub-goals' before reaching their destination. For our purposes, we can split the eye and mouse time-series data into these phases to better identify unconscious points of friction.

Importantly, this rich time series data contains information not only about where a person looks and how they move, but also can be used to measure the tight coupling that typically exists between the eye and the hand. That is, within each phase of a task, an intuitive notion is that coordination between the eyes and the hands is necessary for effective interaction [Ballard et al. 1992]. What is less intuitive is an effective way to calculate this relationship dynamically. One tool quantifies this eye-mouse coordination by calculating the velocities of both the hands and the eyes (i.e. the mouse cursor and gaze position) and checks 1) if they are going in the same direction (i.e. similar vectors) and 2) if the velocities are close to one another. This method generates what is known as a Tlead value [Deng et al. 2016], which approximates how far ahead the eyes are of the mouse cursor (or vice versa). Here, we can use the Tlead value as a measurement of the relationship that exists between the eyes and the hands dynamically. These types of measures are best used in dynamic environments, which most UX tends to exist in.

Assessing dynamic behaviours

In the following study, we investigated the utility of tracking mouse and gaze position using both a monitor-mounted eye tracker and a webcam eye tracker during naturalistic, video game-based UI menu navigation. We did this in three distinct ways: 1) splitting the task into smaller 'phases' such that each phase could be analyzed individually, 2) assessing the hand-eye coordination relationship dynamically over these phases and, 3) collecting data in two groups: a local and a remote cohort which had either simultaneous hardware and webcam-based eye tracking (local) or only webcam (remote)

We predicted that user friction could be detected through eye and hand movements when a UI interaction was broken into task relevant phases. Additionally, we were interested in the limitations of webcam-based eye tracking, and whether its shortcomings would prevent its effective use as a UXR tool. To address these questions, we collected data from two cohorts. The first cohort (Local) was studied in-person using two types of eye trackers (dedicated hardware and webcam), whereas the second cohort (Remote) completed the task remotely through the Amazon MTurk service using webcam eye tracking only. At the beginning of each trial, participants were given a task to complete through a simple text prompt. We experimentally induced friction by cuing participants about their specific goal using either a Direct prompt or an Indirect prompt. Direct prompts were explicit about which menu items to interact with and in what order, while Indirect prompts were more vague, presenting the task at a much higher level. We predicted that the Indirect prompts would be harder for the participants to complete, resulting in longer completion times and more mouse and gaze movement. Our goal was that this intentional manipulation would validate our measures such that any *other* changes could be taken as signs of naturally occurring friction. To address our second aim of determining the utility of webcam eye trackers, the dedicated hardware and webcam eye trackers were directly compared in the Local cohort to get a sense of the overall accuracy and performance of each. We predicted that even in the face of reduced temporal and spatial accuracy, webcam eye trackers would be capable of detecting any friction points identified using the higher-quality eye tracker. Finally, the Remote cohort allowed us to test if the effects found in the Local cohort could be replicated using only a webcam and where we did not control the collection environment. We predicted that we would see similar friction points in both the Remote and Local cohorts, while also not introducing a significant amount of noise into the data.

METHODS

Participants

Ethical approval was granted by the University of Alberta Health Research Ethics Board under protocol Pro00087329 and ethical protocols were in adherence to the 1964 Declaration of Helsinki.

Ten BioWare employees were recruited for the local-cohort (all male, mean age: 30.6 ± 5.2 years). All participants gave informed consent to participate in the current study. Three participants had to be removed from the data pool (one had unusable data, and two had recording errors). We acknowledge that this is a gender biased sample, but it was a convenience sample and likely reflects the actual gender imbalance in this industry.

Thirty-eight subjects participated in the remote-cohort using the Amazon MTurk and Prolific platforms (19 females, mean age: 30.6 ± 9.5 years). All remote subjects gave informed consent to participate in the current study. No participants were removed from the dataset. All remote participants were paid $7.50 USD for their participation in this study.

Equipment

For the local cohort, two different eye trackers were used. The first was a Tobii Eye Tracker 4C: a consumer-grade monitor-mounted eye tracking solution, which offers data collection at 90 Hz at a cost of around $200 CAD. The second was a consumer-grade webcam (Logitech C270), capable of collecting data at a resolution of 1280×720 pixels at 30 frames per second. The purpose of this webcam was not to find the most powerful or feature-rich device, but rather to use a webcam that a typical person may own. Additionally, a standard optical wired desktop mouse was used (Dell MS116).

For the remote cohort, subjects required their own computer with a physical mouse (instead of a trackpad) and a webcam. We did not discriminate on the quality of the webcam, as we were interested to collect data from a wide range of computer setups. As such, we do not know the average resolution or collection frequency (i.e. frames per second) of the webcams used in the remote portion of data collection. What we can report is that the average sampling rate of the webcam eye tracker algorithm [Finger et al. 2017] was 15.17Hz (± 8.56Hz), which reflects not just the capabilities of the webcam, but also a participant's computer specifications and internet connection speed.

Software. For the Local cohort, custom software was written in C# to access and collect data from the Tobii eye tracker using the official Tobii SDK. Additional programs were also written to access and collect data about the mouse position and mouse clicks over time, again in C#[1]. Because there are multiple data streams that are being recorded simultaneously, it is important to ensure that all of these data streams are synchronized using a common source for time-stamping. We used Lab Streaming Layer [SCCN 2021], a data stream synchronization library designed specifically for this purpose.

For the webcam-based eye tracker (Local and Remote cohorts), we used an implementation built by Labvanced [Finger et al. 2017] that is capable of tracking gaze positions on the screen in their platform. At the beginning of each trial, a short (30s) calibration task appeared on the screen where the participant was guided to fixate on points in a circle. Through the use of websockets, a custom program written in Python triggered the recording of gaze positions on the screen, giving access to the timing information so we could later synchronize the webcam gaze data back to the mouse and other gaze (Tobii) data (Local cohort only).

For the Remote cohort, we relied on the built-in Labvanced mouse tracker, which sampled data at 60hz and was automatically synchronized with the Labvanced eye tracking data.

Task

The menu navigation task used for this study was designed in Labvanced. Labvanced is an online site-as-a-service that allows users to create and deploy psychology experiments [Finger et al. 2017]. Because we were interested in replicating an authentic user experience, we recreated the main menu layout from a popular video game created by BioWare: Mass Effect 3 (ME3; [BioWare 2012]). The Labvanced task is an abstract version of the menus displayed in ME3 without any of the complex visual elements present in the original game (e.g such as the bright colours, or starry backgrounds; see Figure 1). We wanted users to navigate to common menu destinations, where typical users make adjustments such as to the game audio, in-game settings, or graphical settings. We replicated the following menu options, herein referred to as *goal frames*: Accomplishments, Gameplay, Graphics, Mouse, Narrative, and Sound.

The general procedure of the task is as follows: the participant is given a prompt where they are given a goal to complete. Once acknowledged, the participant then navigates through the menu to the goal frame where they perform a target interaction (e.g. check a box, find a piece of information, adjust a scroll bar, etc), before navigating back to the main menu to click on *"Exit game"*. This will trigger the beginning of the next trial. For an example visual of the general flow of the task, see Figure 1. In the example, the user is told to navigate to the *Graphics* goal frame and to turn on 'Antialiasing'.

Task design. Two types of prompts were presented to the participant throughout the task: a *Direct* or *Indirect* prompt.

A Direct prompt is a clear and concise set of instructions that guides the user towards a specific end goal by providing the names of intermediary menu buttons that they need to click on or adjust in order to reach the target.. An example of an Direct prompt is: *"Go to Extras - Options - Gameplay and turn Hints on."*. This prompt is meant to give direct instructions about which buttons to click on, while taking away any extra information.

The Indirect prompt aims to mimic the thought process of a video game player who wants to access the settings menu to make necessary adjustments. For example, one prompt might be *"You notice you are having trouble hearing the characters' dialogue in game. Go and turn on the subtitles."*. These prompts are necessarily more abstract in that they give less direct information to the player, but enough information to complete the task. As previously discussed, this manipulation was

[1]All of the software used to collect the data are open source and available at the following URLs: https://github.com/scottastone/TobiiGazeLSL and https://github.com/scottastone/MouseLSLGUI

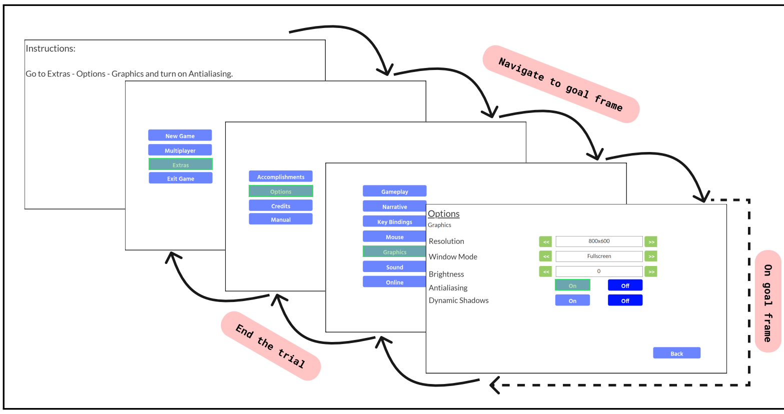

Fig. 1. A waterfall representation of the order of frames that the participant will encounter in a typical trial. At the top left is the first frame of the trial the user will see, known as the prompt. The prompt contains all of the instructions necessary to complete the trial. Each frame contains a goal target, which is highlighted in green in this figure only for clarity. Upon clicking continue (bottom centre of the left-most frame; not shown), the participant will enter the next frame (moving clockwise), starting the trial. The user continues to navigate through the menu until they reach the goal frame (bottom right frame). This is the *"Navigate to the goal frame"* phase. The participant then completes the given task on the goal frame (*"On goal frame"*). Next, the participant will work to *"End the trial"* by moving backwards (moving counter-clockwise) in the menu towards the first frame they encountered following the prompt. Here, they must click "Exit game" to start the next trial.

introduced specifically because we expected the Indirect prompts to be more difficult for the end user. In general, if any friction occurred during the task, the Indirect prompt should exacerbate that friction.

This task design was chosen to allow the participant to more fully explore the task space without being explicitly told to do so. An Indirect prompt does exactly this. Some of this time spent exploring will not be fruitful, which we interpret as unwanted friction. Conversely, we expected Direct prompts to have less exploration time, as the interim buttons needed to reach the goal frame are given directly. However, some goal frames will require lots of exploration (such as Accomplishments) because providing information is their primary role, rather than being used to make adjustments (i.e. Graphics goal frame). The key difference here then, is understanding the context in which the goal frames *should* be used in order to detect friction within them.

Procedure

For the Local cohort, the participant was brought into the room and seated in a comfortable computer chair. They were given the opportunity to adjust the height, lumbar, and arm heights to ensure they were comfortable for the duration of the experiment. They were then positioned in front of the computer monitor (27" Dell, 2560x1440 pixels, 60Hz), approximately 40cm away. If the participant was wearing a mask, they were asked to remove it, as the webcam eye tracking algorithm does not work with one on. Each participant was given up to 10 practice trials on the task to familiarize themselves with the procedure, though none of the participants needed all 10. Upon ensuring that the participant was comfortable and understood the task, the experimenter began the study and left the room. The participant was given a total of 96 trials (48 Direct prompts and 48 Indirect prompts). The task took approximately one hour to complete.

Remote cohort subjects received on-screen instructions to seat themselves comfortably and ensure they had ample lighting in the room. Collecting data remotely meant we had no guarantees of environment or posture, nor could we know what hardware was being used by the participant (e.g. monitor size and resolution were not collected). Each participant was given up to 10 practice trials on the task to familiarize themselves with the general flow of the study. After completing webcam calibration, the participant was given a total of 104 trials (52 Direct prompts and 52 Indirect prompts). The task took approximately one hour to complete. For the remote cohort only, we re-balanced the task to include more equal representation of each goal frame. As such, instead of having a total of 96 trials, participants completed a total of 104 trials[2]. The basic presentation of the prompt and overall flow of the task however remained identical and took the same amount of time.

DEPENDENT VARIABLES

There are four general categories of data collected: *time*, *mouse*, *gaze*, and *coordination*. Within each category subdivided the data into three measures defined by distinct phases: navigating to the goal frame, on the goal frame, and ending the trial (see Figure 1 to see how the phases were split). Please note that *mouse* data were not analyzed, as its dynamics are captured by the *coordination* measure.

Time. All time data are reported in seconds, split into each of the three phases.

Gaze. Gaze distance (e.g. the cumulative amount of movement) in each phase was calculated. To test for the predicted accuracy drop off in webcam data we calculated two additional gaze measures for the interactions on the goal frame: time in goal target and minimum distance to goal target. Time in goal target is the amount of time the gaze was within the bounds of the goal target, which was the piece of information that needed to be interacted with on the goal frame. Minimum distance is the smallest distance between any gaze sample and the bounds of the goal target, with this value achieving 0 if the gaze was within the goal target at any time. A lower accuracy system should have a larger minimum distance and less time in target. All gaze data are reported in standardized units and was sampled at 90 hz (Tobii data) and approximately 20 hz (Webcam data), which were up-sampled to 90hz for analysis. The units were standardized by converting all pixel coordinates to fit the Labvanced coordinate space of 800×450 units.

Coordination. The amount of time the eyes and hands moved together was quantified using Tlead [Deng et al. 2016] and examined across each phase. Tlead calculations necessitate data that are sampled at identical frequencies, meaning the gaze and mouse data must be re-sampled to a common sampling rate. While the calculation of Tlead returns three possible values: NaN proportion (meaning one or both of the data streams are decoupled from one another), positive proportion (gaze leading the cursor), and negative proportion (cursor leading gaze), we were primarily interested in the amount of coupling observed. Therefore, for our analysis, we collapsed Tlead into the percent of time the eyes and hands were either coupled (e.g. signed; non-NaN) or decoupled (NaN). Importantly, not all interactions require tight eye-hand coupling; sometimes our eyes collect information in one space, but our hands work in another. Our novel use of Tlead allows us to quantify tasks that require tight coupling versus those that force eye-hand decoupling.

[2]All of the prompts can be found at the data repository: https://osf.io/f49xg/

RESULTS

Statistical test designs

All statistics were calculated using Jamovi 2.3.18 [jamovi 2022]. Repeated measures ANOVA (rmANOVA) were sphericity-corrected using the Greenhouse-Geisser estimate of the F statistic where necessary.

For the Local cohort, because two different eye trackers were used concurrently, tests that investigate gaze-based measures have a 6×2×2 rmANOVA design: 6 GoalFrames (accomplishments, gameplay, graphics, mouse, narrative, sound), 2 Conditions (Direct, Indirect) and 2 Eyetrackers (Webcam and Monitor-mounted). Mouse and time measures were identical regardless of the eye tracker used, so the rmANOVA design only required a 6 × 2 design (GoalFrame × Condition). To test coordination measures, we used an rmANOVA with a 6 × 2 × 2 design (GoalFrame × Condition × Eyetracker).

Since we were interested in comparing the performance of the Remote and Local cohorts, we also used a mixed ANOVA with a 6 × 2 × 2 design (GoalFrame × Condition) for within subjects and the subject Cohort (Local-webcam, Remote) as the between subjects factor to compare across all of our measures. For these tests, only results with significant main effects of Cohort or interactions involving Cohort will be reported.

Statistics: Time

Time: to navigate to the goal frame. These data are the average time that it took a participant to locate the intended goal frame given by the prompt. In the Local cohort, a main effect of GoalFrame was detected (F(1,1.411) = 15.274, p = 0.003, η^2 = 0.403), with Narrative taking the longest for participants to find. The Remote cohort showed the same pattern with no interactions involving Cohort. A main effect of Cohort was detected (F(1,1) = 4.503, p = 0.039, η^2 = 0.041), where the Remote cohort took longer, which we interpret to be an expertise effect between the cohorts.

Time: on the goal frame. This is the average amount of time the participant spent to complete the task on the goal frame given by the prompt. In the Local cohort, a main effect of Condition was detected (F(1,1) = 35.743, p < 0.001, η^2 = 0.014), with Indirect prompts taking longer than Direct prompts. This suggests that Indirect prompts do in fact take longer to complete, agreeing with our earlier prediction. A main effect of GoalFrame was detected (F(1,1.791) = 160.472, p < 0.001, η^2 = 0.898), with Accomplishments taking the longest time to complete the task, followed by Sound. A significant GoalFrame × Condition interaction was detected (F(1,13.104) = 10.236, p = 0.002, η^2 = 0.024), where users tended to take longer when given an Indirect prompt on all goal frames with the exception of Narrative. The Remote cohort showed the same pattern and the cohort comparison showed no significant main effects or interactions involving Cohort.

Time: to end trial. This is the average time it takes for the user to end the trial after they make the intended manipulation on the goal frame. In the Local cohort, a main effect of Condition was detected (F(1,1) = 7.889, p = 0.031, η^2 = 0.074), where participants took longer to end the trial on Indirect prompts than Direct prompts. The Remote cohort showed the same pattern and the cohort comparison showed no significant main effects or interactions involving Cohort.

Statistics: Gaze

Gaze: distance to navigate to the goal frame. This is the average distance in pixels that the gaze traveled on the screen when the participant was navigating to the goal frame. See Figure 3A. In the Local cohort, a significant main effect of GoalFrame was detected (F(1,1.530) = 16.732, p < 0.001, η^2 = 0.306), with Narrative goal frames requiring the most gaze distance to find. A significant main

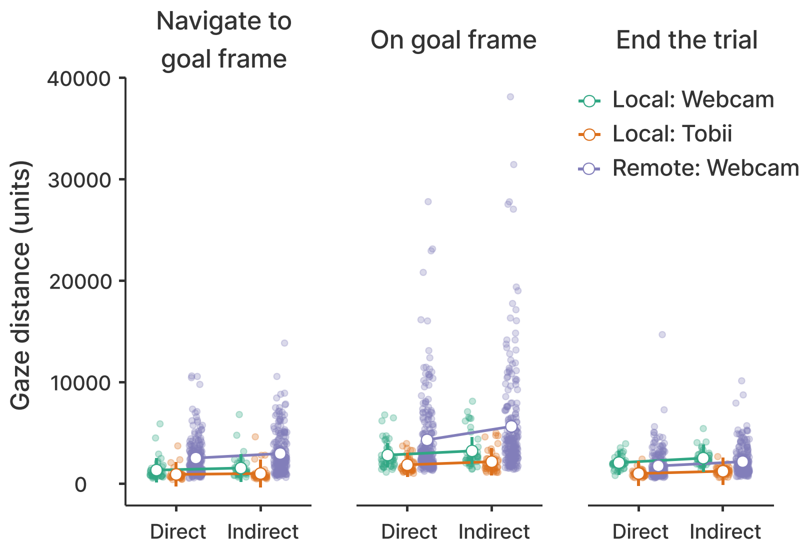

Fig. 2. The eye distance traveled over the course of an entire trial split by the phase. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. The green circles and lines are webcam eye tracker data from the Local cohort, the orange circles and lines are the Tobii eye tracker data from the Local cohort, and the purple circles and lines are the webcam eye tracker data from the Remote cohort. The X axis shows which kind of prompt was presented to the participant. The Y axis shows the average distance traveled in units. Each data point is scatter-plotted under each mean. 95% confidence intervals are plotted around the means. A significant main effect of Condition demonstrates that Indirect prompts result in more distance being traveled. A significant interaction between Condition × Eyetracker shows the webcam accumulates more error over time.

effect of Eyetracker was detected (F(1,1) = 8.819, p = 0.025, η^2 = 0.067), with the webcam eye tracker overestimating the distance traveled relative the monitor-mounted system. A significant GoalFrame × Eyetracker interaction was detected (F(1,1.983) = 6.539, p = 0.012, η^2 = 0.001), where goal frames that required more time to complete also resulted in higher estimates for distance traveled for webcam eye tracker data relative to the monitor-mounted eye tracker. This suggests the webcam eye tracker likely overestimates the distance traveled and this effect scales with time. The Remote cohort showed the same pattern with no interactions involving Cohort. A main effect of Cohort was detected (F(1,1) = 5.233, p = 0.027, η^2 = 0.069), where Remote participants moved their eyes more, likely due to the expertise differences between the cohorts.

Gaze: distance on the goal frame. This is the average distance in pixels that the gaze traveled on the screen when the participant was on the intended goal frame completing the task outlined by the given prompt. See Figure 3B. In the Local cohort, a significant main effect of Condition was detected (F(1,1) = 17.499, p = 0.006, η^2 = 0.016), where Indirect prompts required more eye movements than Direct prompts. Again, this supports our earlier prediction that Indirect prompts are more difficult, and will thus take longer to complete with more eye movements. A significant main effect of GoalFrame was detected (F(1,2.151) = 60.987, p < 0.001, η^2 = 0.566), with Accomplishments requiring the most gaze movements overall. A significant main effect of Eyetracker was detected (F(1,1) = 16.356, p = 0.007, η^2 = 0.134), where the webcam eye tracker overestimated the distance traveled relative to the monitor-mounted eye tracker. A significant GoalFrame × Condition interaction was detected (F(1,2.482) = 5.090, p = 0.016, η^2 = 0.034), where Indirect prompts resulted in higher distances traveled, with the exception of on Mouse and Narrative goal frames. A significant GoalFrame × Eyetracker interaction was detected ($F(1,1.251) = 14.540$, $p = 0.004$, $\eta^2 = 0.039$), where goal frames that took longer to complete had disproportionately higher distances traveled for the webcam eye tracker versus the monitor-mounted eye tracker. A Condition × Eyetracker interaction was detected ($F(1,1) = 11.591$, $p = 0.014$, $\eta^2 = 0.000$), where the webcam eye tracker resulted in a higher travel distance difference (compared to the monitor-mounted eye tracker) on Indirect prompts. For the cohort comparison, a significant main effect of Cohort was detected ($F(1,1) = 4.857$, $p = 0.033$, $\eta^2 = 0.041$), where Remote participants had more gaze movements on the goal frame. A significant GoalFrame × Cohort interaction was detected ($F(1,1.142) = 7.607$, $p = 0.006$, $\eta^2 = 0.046$), where Remote participants had more gaze movements on the Accomplishments goal frame. A significant Condition × Cohort interaction was detected ($F(1,1) = 9.685$, $p = 0.003$, $\eta^2 = 0.002$), where Remote participants had more gaze movements when given a Indirect prompt, again likely attributed to the Remote cohort's inexperience with the menu layout. A visual summary of these results can be seen in Figure 3B.

Gaze: distance to end trial. This is the average distance in pixels that the gaze traveled on the screen when the participant had completed the task outlined by the prompt and was on their way to end the current trial. In the Local cohort, a significant main effect of Condition was detected ($F(1,1) = 20.480$, $p = 0.004$, $\eta^2 = 0.028$), where Indirect prompts resulted in more gaze distance traveled. A significant main effect of Eyetracker was detected ($F(1,1) = 25.193$, $p = 0.002$, $\eta^2 = 0.504$), where the webcam eye tracker traveled a further distance. A significant Condition × Eyetracker interaction was detected ($F(1,1) = 9.421$, $p = 0.022$, $\eta^2 = 0.004$), where the web cam eye tracker traveled more during the Indirect prompts. The Remote cohort showed the same pattern and the cohort comparison showed no significant main effects or interactions involving Cohort.

Gaze: minimum distance to goal target. This is the minimum distance between the gaze and the goal target(s) on the goal frame throughout an entire trial. A significant main effect of Eyetracker was detected ($F(1,1) = 21.146$, $p = 0.004$, $\eta^2 = 0.486$), where the webcam eye tracker had higher distances to any of the goal targets on screen, confirming the webcam eye tracker was overall less spatially accurate. A main effect of Cohort was detected ($F(1,1) = 7.589$, $p = 0.009$, $\eta^2 = 0.092$), where the Remote cohort had a lower minimum distance. A significant Cohort × Condition interaction was detected ($F(1,1) = 4.907$, $p = 0.032$, $\eta^2 = 0.006$) where the Local cohort had a larger decrease in minimum distance between conditions than the Remote cohort.

Gaze: on target time. This is the average amount of time the gaze data were within the bounds of any of the intended targets on the goal frame outlined by the prompt. A value of 0 indicates that gaze was never within the bounds of the goal targets on the goal frame. In the Local cohort, a significant main effect of GoalFrame was detected ($F(1,2.317) = 2.317$, $p < 0.001$, $\eta^2 = 0.073$), where gaze was within the bounds of the goal targets on the Narrative goal frame. A significant main effect of Condition was detected ($F(1,1) = 7.342$, $p = 0.035$, $\eta^2 = 0.005$), where Indirect prompts lead to a longer dwell time on the goal targets. A significant main effect of Eyetracker was detected ($F(1,1) = 69.095$, $p < 0.001$, $\eta^2 = 0.663$), where the webcam eye tracker had significantly less time spent within the target bounds. A significant GoalFrame × Eyetracker interaction was detected ($F(1,2.324) = 13.451$, $p < 0.001$, $\eta^2 = 0.064$), where the time spent on the Narrative goal frame using the monitor-mounted eye tracker had disproportionately more time spent within the target goal target bounds than other goal frames. A significant Condition × Eyetracker interaction was detected ($F(1,1) = 6.840$, $p = 0.040$, $\eta^2 = 0.004$) where the amount of time gaze data were within the bounds of the goal target was equal across Condition, whereas the monitor-mounted eye tracker had higher values for Indirect prompts. For the cohort comparison, a main effect of Cohort was detected ($F(1,1) = 4.269$, $p = 0.045$, $\eta^2 =$

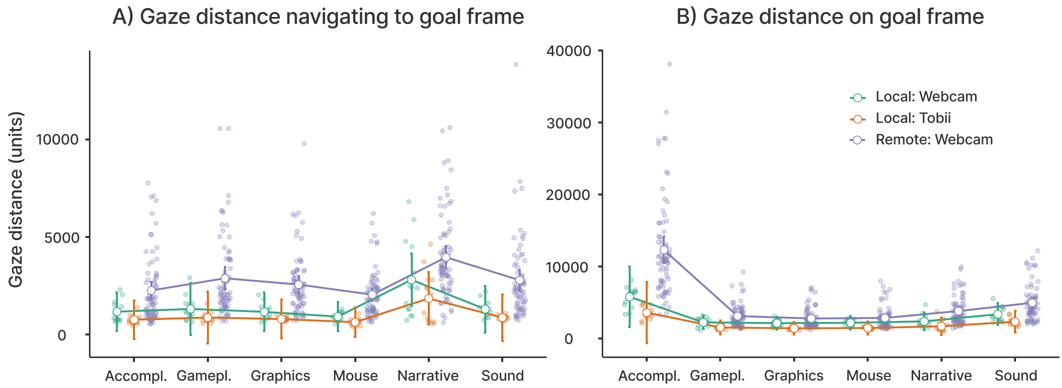

Fig. 3. Plots of the gaze distance in two phases: navigation and on the goal frame. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. The green circles and lines are webcam eye tracker data from the Local cohort, the orange circles and lines are the Tobii eye tracker data from the Local cohort, and the purple circles and lines are the webcam eye tracker data from the Remote cohort. The X axis has each of the goal frames (with Accompl. and Gamepl. for Accomplishments and Gameplay, respectively). The Y axis is the distance traveled in standardized units. The underlying data are scatter-plotted under each mean. 95% confidence intervals are plotted around the means. A) The average gaze distance traveled navigating to the goal frame. Here, we can see an increase in the amount of gaze movement required to enter the Narrative goal frame, regardless of the participant pool or eye tracker used. B) The average gaze distance traveled while on the intended goal frame. Here, we can see that Accomplishments required more gaze movements relative to other frames. The Remote cohort looked around the most, which is indicative of their relative inexperience with the menu and variable collection environment.

0.036), where the Remote cohort looked at targets for longer, likely reflecting their inexperience with the menu system. A significant GoalFrame × Cohort interaction was detected (F(1,3.755) = 2.491, p = 0.049, η^2 = 0.018), where the Remote cohort looked at targets on most goal frames more, with the exception of Narrative and Sound.

Statistics: Coordination

Coordination: Tlead. This is the percent of time the eyes and hands were uncoupled. In the Local cohort, a significant main effect of GoalFrame was detected (F(1,2.059) = 241.902, p < 0.001, η^2 = 0.762), where the the eyes and hands were disproportionately uncoupled on the Accomplishments goal frame relative to the others. A significant main effect of Eyetracker was detected (F(1,1) = 209.679, p < 0.001, η^2 = 0.016), where data collected with the Webcam were more coupled. A significant GoalFrame × Eyetracker interaction was detected (F(1,2.144) = 6.413, p = 0.011, η^2 = 0.000), where the Webcam eye tracker had a larger difference between coupled and uncoupled proportions for all goal frames. See Figure 4 for a visual of the results. For the cohort comparison, no main effect of Cohort was detected. A significant GoalFrame × Cohort interaction was detected ((F(1,2.521) = 3.881, p = 0.016, η^2 = 0.006), which seems to be driven by the Local cohort showing slightly more coupling on the Narrative goal frame. A significant Condition × Cohort interaction was detected ((F(1,1) = 16.663, p < 0.001, η^2 = 0.003), where the Remote cohort had a slightly larger difference between the prompts given.

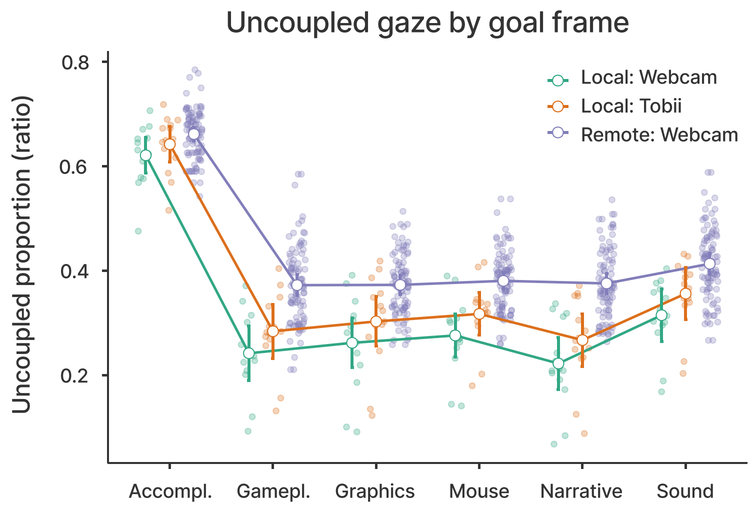

Fig. 4. The average proportion of $Tlead_{NaN}$ values for each type of goal frame. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. The green circles and lines are webcam eye tracker data from the Local cohort, the orange circles and lines are the Tobii eye tracker data from the Local cohort, and the purple circles and lines are the webcam eye tracker data from the Remote cohort. The X axis has each of the goal frames (with Accompl. and Gamepl. for Accomplishments and Gameplay, respectively). The Y axis is proportion of $Tlead_{NaN}$ values as a ratio. The underlying data are scatter-plotted under each mean. 95% confidence intervals are plotted around the means. Here, we can see that there are significantly more $Tlead_{NaN}$ values on the Accomplishments goal frame, suggesting that the eyes and hands are dissociated from one another for this task. The eyes are drawn to search on one part of the screen while the hands click relatively stationary in another.

GENERAL DISCUSSION

Main findings

We collected eye, hand and coordination measures in two cohorts of participants as they completed a simple UI navigation task. For the Local cohort, we investigated a consumer-grade monitor-mounted eye tracker and directly compared it to a simultaneously recorded webcam-based eye tracking algorithm. Because both datasets were recorded concurrently in the Local cohort, we were able to directly compare performance. To our surprise, we found that the webcam produced sufficient data quality to be directly comparable to the more sophisticated monitor-mounted eye tracker in revealing important features of a user's experience. While the temporal and spatial resolution was much lower in the webcam data, we were still able to detect a friction point in the menu design that has not, to our knowledge, been discovered previously. For the Remote cohort we looked exclusively at webcam data and did a between groups comparison to the Local data to investigate if our findings would hold once the experimental control over hardware and environment was removed.

In general, we found similar results across both cohorts. In terms of the initial prompt used, an Indirectly worded prompt resulted in worse performance than a Direct one (see Figure 2). That is, users took longer and moved both their mouse and eyes more with an Indirect prompt. In both cohorts, we also found two candidate points of friction. The first candidate friction point was on the Accomplishments goal frame, where participants spent a significant amount of time locating given accomplishments. The second candidate friction point was navigating to the Narrative goal frame.

A conventional analysis using only time-based measures would have concluded that these two delays were both indicators of friction. However, our unique analysis of eye and mouse movements, and especially their coordination, paints quite a different picture. For the Accomplishments frame, we saw an increase in eye-hand decoupling (see Figure 4), indicating much more eye movement than mouse movement. In context of the task given, this was in fact a reasonable result; the Accomplishments task required repeatedly clicking a button with the mouse while reading and searching through text descriptions that appeared with each click. This required lots of gaze movements and few mouse movements, suggesting that the eyes and hands had distinct roles. As a result, this is actually the expected behaviour rather than a point of friction. By comparison, the delay when navigating to the Narrative frame was accompanied by more eye and hand movements, which remained coupled. The extended search time and increased eye and hand movements indicate confused and inefficient exploration (eye) and exploitation (hand). Only the collection and analysis of gaze and movement behaviours was able to disambiguate these cases and identify a true point of friction.

Local versus Remote cohorts

While the level of congruence between the two cohorts was remarkable, especially considering the lack of experimental control over the conditions in which the Remote cohort was tested, one key difference between cohorts was the level of experience. The Local cohort were all employees of the company that developed the video game the UI task was based on, whereas the Remote cohort did not necessarily have any experience with video games. Collecting data from less-experienced users can help pinpoint potential issues that the more experienced group may have simply adapted to over time. For example, a new video game player may not be familiar with game-specific terminologies used in the menus (e.g. graphical options such as vertical synchronization, or mouse sensitivity adjustments). Our findings support this conclusion, showing that the Remote cohort was significantly slower and had more gaze and hand movements than the Local cohort, but *only* when they were operating on the goal frame (see Figure 3). That is, they exhibited no difference in the mechanics of the task (moving the mouse, clicking buttons, exiting each trial) and their inexperience was only evident when task relevant knowledge was expected to play a significant factor. Interestingly, regardless of the level of experience, the friction point described above naturally emerged in the data for users in both cohorts. As researchers, we are able to pinpoint specifically when the friction was occurring, and provide actionable insight into how to fix the problem. Here, we saw that all users tended to have issues finding the Narrative goal frame, and one potential reason for this could be that the name of the menu is ambiguous or may not reflect its contents for most users. A UX researcher could test this hypothesis by altering the menu's name and testing if the friction still exists. It is also possible to imagine expertise-dependent adjustments UX designers may wish to use, such as tailoring an experience better suited for a novice as compared to an expert user. Additionally, it is important for the UX researcher to contextualize the utility of friction. Some friction can actually be useful to ensure a user is paying attention to critical components and can lead to higher user satisfaction [Mejtoft et al. 2019].

Recording gaze and mouse data gives insights into the end-user's behaviour, most of which is unconsciously controlled [Bridgeman 1992; Freeman et al. 2011; Goodale 2011; Goodale and Milner 1992; Magnuson 2005; Milner and Goodale 2006]. Typically, when assessing user interfaces, researchers use qualitative approaches such as interview-style questions aimed at probing the end-user's conscious experience [Nunnally and Farkas 2016]. We argue that while a qualitative approach might lend to some insights about the overall experience, some friction points cannot be uncovered simply because users themselves may not be aware they were experiencing friction. Here, we argue that the methods employed in the present study demonstrate that friction can be detected through the use of easy-to-deploy hardware and software at a minimal cost. It is important to note that this study did not exhaustively compare the performance of hardware and software eye trackers (for a study investigating this, please see Wisiecka et al. [2022]), so it is important that the researcher knows the inherent limitations of the implementation they choose. Additionally, UX researchers can augment their current approaches by adding our methodology at little (i.e. monitor-mounted eye tracker) to no (i.e. webcam eye tracker) cost, both financially and methodologically.

The webcam eye tracker was less accurate and had a lower temporal resolution than the dedicated eye tracker, but this did not impede data analysis or interpretation for our design. However, if high accuracy or temporal resolution is a necessity for the experimental design, webcam eye tracking may not be a suitable choice. For example, researchers interested in speed-accuracy trade-offs would likely benefit from higher-powered systems. When looking at the minimum gaze distance to any of the targets, we found that the webcam was consistently about 20 pixels (or 6.25 units[3]) away from the goal target in the Local cohort. An offset or threshold can boost accuracy, but target size matters. Improved eye-tracking algorithms may reduce the need for dedicated systems in future studies as spatial and temporal accuracy increases.

Moving beyond the laboratory

In general, we look at the data in the Remote group as an exercise in a cognitive ethology-based approach towards data collection [Kingstone et al. 2008]. Cognitive ethology calls for moving beyond some laboratory-based assumptions (e.g. cognitive process invariance and control) to better understand the naturalistic variance that occurs in real-world complex tasks [Chisholm et al. 2014; Kingstone et al. 2005, 2008; Smilek et al. 2006]. In the Remote cohort, the noise in the collection environment did not appear to overwhelm the signal as evidenced by the similar findings between the studies. Arguably, some scientific findings lose sight of the real-world implications of their research; the environment in which the data were collected is not similar to how people actually behave. This approach challenges many of the control assumptions we as scientists make when conducting our studies. Is it critical that the users sit exactly 40cm from the screen? Does the lighting of the room need to be identical between participants? Our results suggest that, at least in this case, that may not be necessary. It is hard to know what the trade-off between data quality and real-life applicability truly is unless we are conducting ecologically valid studies in the first place. Perhaps it is reasonable to sacrifice data quality for the sake of having a closer understanding of the cognitive processes in the environment they are naturally practiced. This is particularly true for UX researchers interested in getting a deeper insight into how and where users *actually* use and encounter their products.

Privacy

Naturally, as we move towards more seamless data collection (i.e. the participant may not even realize their gaze is being tracked), privacy concerns arise. Many webcam eye tracking algorithms are collected and processed on the local machine (e.g. WebGazer [Papoutsaki 2015; Yang and Krajbich 2020] and Labvanced [Finger et al. 2017]) or obtains the user's consent prior to recording [Finger et al. 2017] and thus require the end-user to be aware of its presence. The current iteration of the algorithm used in the present task also required inter-stimulus calibration periods, making it obvious to the end-user that their gaze was being recorded. However, in the future it is likely that gaze tracking models will become advanced enough to not require frequent re-calibrations

[3]The collection environment was 800x450 units, and we can extrapolate pixels if we know the monitor resolution, which we did for the Local cohort.

and as a result may become functionally invisible to the end user. We believe that the end-user should always be made aware of (and asked for consent for) the recording of their data, as eye movement patterns can be used to identify unique individuals [Holland and Komogortsev 2011]. This is a major concern that will likely require legislation to capably handle.

Conclusion

Many qualitative methods are used simply because they make logical sense; who else would know better about the user's experience than the user themselves? We provide evidence that there's more than one way to find friction in a design, and it is not mutually exclusive to the current methods used. Eye and mouse tracking can provide a wealth of knowledge to the UX researcher at very little cost. The presented methods may currently be suitable for UX researchers, with the caveat that there is no standardized (or easy) way to analyze the data. Analysis requires intimate knowledge of data cleaning methods from eye and mouse trackers. Future use of eye and mouse tracking in UXR should include efforts to, ironically enough, improve the experience of the UX researchers through standardized tools for cleaning and analysis.

Acknowledgements

Thanks to everyone at BioWare for their enthusiasm and support. Thanks to Mitacs for the financial support.

REFERENCES

Dana H Ballard, Mary M Hayhoe, Feng Li, and Steven D Whitehead. 1992. Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 337, 1281 (1992), 331–339.

BioWare. 2012. Mass Effect 3. https://www.bioware.com/games/#mass-effect-3

Bruce Bridgeman. 1992. Conscious vs Unconscious Processes: The Case of Vision. *Theory Psychol.* 2, 1 (Feb. 1992), 73–88. https://doi.org/10.1177/0959354392021004

Craig S Chapman, Jason P Gallivan, Daniel K Wood, Jennifer L Milne, Jody C Culham, and Melvyn A Goodale. 2010. Reaching for the unknown: multiple target encoding and real-time decision-making in a rapid reach task. *Cognition* 116, 2 (Aug. 2010), 168–176. https://doi.org/10.1016/j.cognition.2010.04.008

Joseph D Chisholm, Craig S Chapman, Marvin Amm, Walter F Bischof, Dan Smilek, and Alan Kingstone. 2014. A Cognitive Ethology Study of First- and Third-Person Perspectives. , e92696 pages. https://doi.org/10.1371/journal.pone.0092696

Shujie Deng, Jian Chang, Julie A Kirkby, and Jian J Zhang. 2016. Gaze–mouse coordinated movements and dependency with coordination demands in tracing. *Behav. Inf. Technol.* 35, 8 (Aug. 2016), 665–679. https://doi.org/10.1080/0144929X.2016.1181209

Holger Finger, Caspar Goeke, Dorena Diekamp, Kai Standvoß, and Peter König. 2017. LabVanced: a unified JavaScript framework for online studies. In *International Conference on Computational Social Science (Cologne)*. https://www.researchgate.net/profile/Caspar-Goeke/publication/322273524_LabVanced_A_Unified_JavaScript_Framework_for_Online_Studies/links/5a4f7ac64585151ee284d8c2/LabVanced-A-Unified-JavaScript-Framework-for-Online-Studies.pdf

Jonathan B Freeman, Rick Dale, and Thomas A Farmer. 2011. Hand in motion reveals mind in motion. *Front. Psychol.* 2 (April 2011), 59. https://doi.org/10.3389/fpsyg.2011.00059

Jason P Gallivan and Craig S Chapman. 2014. Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets. https://doi.org/10.3389/fnins.2014.00215

Jason P Gallivan, Craig S Chapman, Daniel M Wolpert, and J Randall Flanagan. 2018. Decision-making in sensorimotor control. *Nature Reviews Neuroscience* 19, 9 (2018), 519–534.

Melvyn A Goodale. 2011. Transforming vision into action. *Vision Res.* 51, 13 (July 2011), 1567–1587. https://doi.org/10.1016/j.visres.2010.07.027

M A Goodale and A D Milner. 1992. Separate visual pathways for perception and action. *Trends Neurosci.* 15, 1 (Jan. 1992), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

M A Goodale, D Pelisson, and C Prablanc. 1986. Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature* 320, 6064 (1986), 748–750. https://doi.org/10.1038/320748a0

Qi Guo, Haojian Jin, Dmitry Lagun, Shuai Yuan, and Eugene Agichtein. 2013. Mining touch interaction data on mobile devices to predict web search result relevance. In *Proceedings of the 36th international ACM SIGIR conference on Research* *and development in information retrieval* (Dublin, Ireland) *(SIGIR '13)*. Association for Computing Machinery, New York, NY, USA, 153–162. https://doi.org/10.1145/2484028.2484100

David Hauger, Alexandros Paramythis, and Stephan Weibelzahl. 2011. Using Browser Interaction Data to Determine Page Reading Behavior. In *User Modeling, Adaption and Personalization*. Springer Berlin Heidelberg, 147–158. https://doi.org/10.1007/978-3-642-22362-4_13

Mary Hayhoe. 2000. Vision Using Routines: A Functional Account of Vision. *Vis. cogn.* 7, 1-3 (Jan. 2000), 43–64. https://doi.org/10.1080/135062800394676

Mary Hayhoe and Dana Ballard. 2005. Eye movements in natural behavior. *Trends Cogn. Sci.* 9, 4 (April 2005), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Andreas Hinderks, Martin Schrepp, Francisco José Domínguez Mayo, María José Escalona, and Jörg Thomaschewski. 2019. Developing a UX KPI based on the user experience questionnaire. *Comput. Stand. Interfaces* 65 (July 2019), 38–44. https://doi.org/10.1016/j.csi.2019.01.007

Corey Holland and Oleg V Komogortsev. 2011. Biometric identification via eye movement scanpaths in reading. In *2011 International Joint Conference on Biometrics (IJCB)*. 1–8. https://doi.org/10.1109/IJCB.2011.6117536

Jeff Huang, Ryen White, and Georg Buscher. 2012. User see, user point: gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1341–1350. https://doi.org/10.1145/2207676.2208591

jamovi. 2022. jamovi (version 2.2.5)[Computer software]. https://www.jamovi.org/

A Kingstone, D Smilek, E Birmingham, and others. 2005. Cognitive ethology: Giving real life to attention research. *Measuring the mind* (2005), 341–357. https://books.google.com/books?hl=en&lr=&id=qTwVDAAAQBAJ&oi=fnd&pg=PA341&dq=Cognitive+ethology+Giving+real+life+attention+research+Kingstone+Smilek&ots=PfAQeFSBX5&sig=jz8-sBMnA6KfFHKbkNnbE5TMx7Q

Alan Kingstone, Daniel Smilek, and John D Eastwood. 2008. Cognitive Ethology: a new approach for studying human cognition. *Br. J. Psychol.* 99, Pt 3 (Aug. 2008), 317–340. https://doi.org/10.1348/000712607X251243

M F Land and M Hayhoe. 2001. In what ways do eye movements contribute to everyday activities? *Vision Res.* 41, 25-26 (2001), 3559–3565. https://doi.org/10.1016/s0042-6989(01)00102-x

Effie Lai-Chong Law, Paul van Schaik, and Virpi Roto. 2014. Attitudes towards user experience (UX) measurement. *Int. J. Hum. Comput. Stud.* 72, 6 (June 2014), 526–541. https://doi.org/10.1016/j.ijhcs.2013.09.006

James S Magnuson. 2005. Moving hand reveals dynamics of thought. *Proc. Natl. Acad. Sci. U. S. A.* 102, 29 (July 2005), 9995–9996. https://doi.org/10.1073/pnas.0504413102

Thomas Mejtoft, Sarah Hale, and Ulrik Söderström. 2019. Design friction. In *Proceedings of the 31st European Conference on Cognitive Ergonomics*. 41–44.

David Milner and Mel Goodale. 2006. *The Visual Brain in Action*. OUP Oxford. https://play.google.com/store/books/details?id=8CaQDwAAQBAJ

Brad Nunnally and David Farkas. 2016. *UX Research: Practical Techniques for Designing Better Products*. "O'Reilly Media, Inc.". https://play.google.com/store/books/details?id=sfOeDQAAQBAJ

Alexandra Papoutsaki. 2015. Scalable Webcam Eye Tracking by Learning from User Interactions. 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, 219–222. https://doi.org/10.1145/2702613.2702627

Alexandra Papoutsaki, James Laskey, and Jeff Huang. 2017. SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR '17)*. Association for Computing Machinery, New York, NY, USA, 17–26. https://doi.org/10.1145/3020165.3020170

SCCN. 2021. Lab Streaming Layer.

Daniel Smilek, Elina Birmingham, David Cameron, Walter Bischof, and Alan Kingstone. 2006. Cognitive Ethology and exploring attention in real-world scenes. *Brain Res.* 1080, 1 (March 2006), 101–119. https://doi.org/10.1016/j.brainres.2005.12.090

Scott A Stone, Quinn A Boser, T Riley Dawson, Albert H Vette, Jacqueline S Hebert, Patrick M Pilarski, and Craig S Chapman. 2022. Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture. *Behav. Res. Methods* (Sept. 2022). https://doi.org/10.3758/s13428-022-01958-6

David A Westwood and Melvyn A Goodale. 2003. Perceptual illusion and the real-time control of action. *Spat. Vis.* 16, 3-4 (2003), 243–254. https://doi.org/10.1163/156856803322467518

Katarzyna Wisiecka, Krzysztof Krejtz, Izabela Krejtz, Damian Sromek, Adam Cellary, Beata Lewandowska, and Andrew Duchowski. 2022. Comparison of Webcam and Remote Eye Tracking. In *2022 Symposium on Eye Tracking Research and Applications*. 1–7.

Nathan J Wispinski, Jason P Gallivan, and Craig S Chapman. 2018. Models, movements, and minds: bridging the gap between decision making and action. *Ann. N. Y. Acad. Sci.* (Oct. 2018). https://doi.org/10.1111/nyas.13973

Xiaozhi Yang and Ian Krajbich. 2020. Webcam-based online eye-tracking for behavioral research. https://doi.org/10.31234/osf.io/qhme6

Received November 4, 2022; revised November 2022; revised February 2023; accepted March 2023