Putkonen, Aini; Jiang, Yue; Zeng, Jingchun; Tammilehto, Olli; Jokinen, Jussi P.P.; Oulasvirta, Antti

Understanding visual search in graphical user interfaces

Published in:
International Journal of Human Computer Studies

DOI:
10.1016/j.ijhcs.2025.103483

Published: 01/05/2025

Document Version
Publisher's PDF, also known as Version of record

Int. J. Human–Computer Studies 199 (2025) 103483

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Understanding visual search in graphical user interfaces

Aini Putkonen [a,*], Yue Jiang [a], Jingchun Zeng [a], Olli Tammilehto [a], Jussi P.P. Jokinen [b], Antti Oulasvirta [a]

[a] Aalto University, Espoo, Finland
[b] University of Jyväskylä, Jyväskylä, Finland

ARTICLE INFO

Dataset link: https://osf.io/hmg9b/

Keywords:
Eye-tracking
Visual search
Dataset
Graphical user interfaces
Linear mixed-effects models

ABSTRACT

How do we find items within graphical user interfaces (GUIs)? Current understanding of this issue relies on studies using symbol matrices, natural scenes, and other non-GUI stimuli. To understand whether the effects discovered in those environments extend to mobile, desktop, and web interfaces, this paper reports on visual search performance and eye movements with 900 real-world GUIs. In an eye-tracking study, participants (N = 84) were given a cue (textual or image) describing a target to find within a GUI. The study found that the type of GUI, the absence/presence of the target, and cue type affected search time more than visual complexity did. We also compared visual search to free-viewing in GUIs, concluding that these two tasks are distinctly different. Synthesis of the results points to a Guess-Scan-Confirm pattern in visual search: in the first few fixations, gaze is frequently directed toward the top-left corner of the screen, a pattern possibly related to the top-left being a statistically likely location of the target or of information that could aid in finding it; attention then gets more selectively guided, in line with the GUI's structure and the features of the target; and, finally, the user must confirm whether the target has been identified or, instead, that no target is visible. The VSGUI10K eye-tracking dataset (10,282 trials) is released for study and modeling of visual search.

1. Introduction

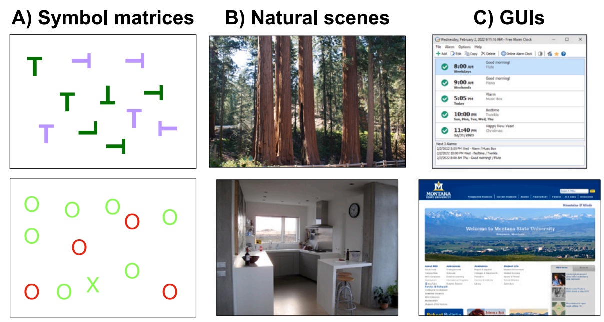
All graphical user interfaces (GUIs) have one thing in common from the user's point of view: ability to interact with them relies critically on finding relevant items in the display area. To send an email or to edit a photo, one must find the pertinent icons, buttons, input fields, labels, and text containers. Failing in this elementary task or spending excessive time on attempting it may have negative consequences for both usability and user experience (Liu et al., 2021; Todi et al., 2019; Ling and van Schaik, 2007). For this reason, understanding and modeling *visual search* – finding targets among distractors – in GUIs has been a foundational topic for human–computer interaction (HCI) research for at least the last three decades (Kieras and Hornof, 2014; Fleetwood and Byrne, 2006; Byrne, 1993; Byrne et al., 1999; Yuan and Li, 2020; Bailly et al., 2014; Todi et al., 2019; Fisher et al., 1989).

Notwithstanding the sustained importance of finding items within GUIs, visual search over *diverse everyday interfaces* has received relatively little empirical attention in recent HCI literature. The reports published on empirical studies of visual search with modern GUIs have focused on specific subtypes of interfaces or not considered eye-tracking data (Jokinen et al., 2020; Bailly et al., 2014; Yuan and Li, 2020; Liu et al., 2021; Byrne, 1993; Jokinen et al., 2017; Pfeuffer and Li, 2018; Halverson and Hornof, 2004; Hornof, 2004; Teo et al., 2012). In fact, much of our knowledge about visual search in HCI is based on findings in controlled studies from other disciplines, which tend to use fairly abstract stimuli, such as symbol matrices (Wolfe, 2021, 2007). A typical study with symbol matrices overlays computer-generated abstract shapes on unicolor backgrounds for tasks such as finding a letter "L" among "T"'s (Wolfe, 2021; Williams, 1967; Wolfe and Horowitz, 2004). It is not clear how well phenomena discovered in these abstract environments generalize to GUIs, which present great variety of visual form and generally lack experimental control. In this paper, we present novel empirical results speaking to that question by releasing a large-scale eye-tracking dataset from visual search within everyday GUIs, VSGUI10K, which richly details the eye movements and visual search times of 84 participants.

Why is it necessary to study visual search in GUIs specifically rather than extrapolate from symbol matrices or other search environments? On one hand, symbol matrices and GUIs share features that make generalizing findings from the former to the latter plausible (see Fig. 1). When compared to symbol matrices, GUIs too are "designed" environments, wherein the placement of each element follows conventions for the visual hierarchy and delineation of objects (Wolfe,

* Corresponding author.
E-mail address: aini.putkonen@aalto.fi (A. Putkonen).

https://doi.org/10.1016/j.ijhcs.2025.103483
Received 26 April 2024; Received in revised form 14 February 2025; Accepted 20 February 2025
Available online 1 March 2025
1071-5819/



		A) Symbol matrices	B) Natural scenes	C) GUIs
I	Natural/ designed	Designed	Natural and/or designed	Designed
II	Everyday stimuli	No	Yes	Yes
III	Delineation of objects	Discrete	Continuous and/or discrete	Discrete
IV	Semantic grouping	No	Yes	Yes
V	Spatial organization	High	Low to medium	High
VI	Variety of visual form	Low	High	Medium to high

Fig. 1. Comparison of symbol matrices, natural scenes, and GUIs as visual search environments. Most work related to visual search focuses on symbol matrices and natural scenes (A-B). Like symbol matrices, GUIs represent a highly designed environment (I) and feature clear delineation of items (III). Natural scenes and GUIs are both everyday stimuli (II), have distinct grouping of elements (e.g., menus and toolbars, IV), feature spatial organization (V), and use a wide range of visual expression mechanisms – colors etc. (VI). Properties IV–VI are regarded as especially important in relation to GUIs' visual complexity (Miniukovich et al., 2018). The illustrative images here are (A) modified from those of Wolfe (2018) and the distractor ratio task (Shen et al., 2000; Myers et al., 2013); (B) reproduced from Flickr under public-domain terms (Creative Commons CC0 1.0) (Levine, 2008; Carroll, 2009); and (C) from the UEyes dataset (Jiang et al., 2023).

1994; Miniukovich et al., 2018), alongside some degree of spatial organization, for instance, via vertical symmetry and the alignment of objects (Miniukovich et al., 2018). Early studies with symbol matrices tended to focus on "feature guidance", or how the characteristics of the target guide search (Wolfe, 2020). Such studies suggest that color and size are among the attributes that guide the deployment of attention in visual search (Wolfe and Horowitz, 2004) and that visual search generally takes longer when none of the targets are present (Chun and Wolfe, 1996) or when more items are shown (Palmer, 1995). There is evidence that these effects are relevant also in the HCI domain (Jokinen et al., 2020; Kieras and Hornof, 2014).

However, even with these similarities, we posit that studying visual search in GUIs specifically is critical for the same reason natural scenes are considered as a separate category in the visual search literature: these environments contain semantic structure that cannot be exploited through feature guidance alone; "scene guidance" must be deployed (Wolfe, 2020). Additionally, work conducted in symbol matrices often probes only a handful of independent variables. This renders it unclear how the results transfer to other contexts (Yuan and Li, 2020), whether GUIs or natural scenes.

Experimental design provides another reason for studying visual search in GUIs. Specifically we draw attention to the importance of understanding visual search in a naturalistic sample of GUIs. Controlled studies in environments that lack semantic structure may not reflect realistic GUI settings. Firstly, such an experimental design may randomize the target's location across all quadrants of the same GUI. This random placement of the target may interfere with scene guidance if the target ends up in an idiosyncratic position. Also, the typical repetition of the target across trials may lead to learning effects, leading to speedier searches in later trials. We can complement such studies via a naturalistic design wherein targets are allowed to vary as they naturally do, in conditions informed by sampling from a large set of everyday GUIs. A design of the latter kind provides the added practical benefit of offering practitioners the ability to predict search times over a wide variety of GUIs. This paper presents an experimental design that balances these considerations. We posit that a successful experimental design representing naturalistic conditions should balance: (1) keeping its stimuli close to those actually encountered by users and (2) introducing control only as necessary to balance the underlying distributions or make data collection feasible.

While experimental designs sensitive to such factors are increasingly common, they prevail mainly in "free-viewing" studies in the domain of visual attention (Shen and Zhao, 2014; Jiang et al., 2023; Leiva et al., 2020; O'Donovan et al., 2014; Borkin et al., 2016). In free-viewing tasks, users are asked to "just look at" the GUI for the prescribed amount of time. It is not obvious how empirical findings from free-viewing should be interpreted in the context of visual search. Indeed, it is well-known in vision research that the way gaze is deployed depends critically on the user's task (Wolfe, 2021). A distinct characteristic of visual search tasks is that a specific target should be found, and the way this is represented in the mind guides search (Wolfe, 2020). In a free-viewing task, though, no such aim is specified. At the same time, the two task types may show similarities. For example, one phenomenon commonly seen in free-viewing of GUIs is a bias wherein the upper-left corner of the display receives a disproportionate number of fixations (Leiva et al., 2020; Jiang et al., 2023). A similar concentration of fixations toward the upper-left is visible in less open-ended tasks, such as browsing of search-engine result pages (Hotchkiss, 2014; Nielsen, 2006). Better assessing such findings' relevance for visual search of GUIs requires moving beyond free-viewing, to understand the distinct tasks' similarities and differences.

In summary, this paper presents four main contributions:

- We test and quantify the effect of several factors on visual search in more ecologically valid circumstances than, to the best of our knowledge, the existing literature contains.
- We compare how visual search compares to free-viewing in GUIs, which so far has been a more popular task in studying behavior in these environments.
- We synthesize our results as a Guess-Scan-Confirm pattern that takes place during visual search.
- We release a novel eye-tracking dataset: `VSGUI10K` covers more than 10,000 visual search trials from 900 real-world webpage, mobile, and desktop interfaces shown to 84 participants.

2. Related work and research questions

2.1. Visual search with symbol matrices and natural scenes

Most work in visual search involving laboratory studies use symbol matrices where the stimuli contain shapes overlaid on often unicolor backgrounds. Attention to natural scenes began growing since differences in the respective environments' characteristics left it unclear whether findings from symbol matrices transfer directly to them (Wolfe et al., 2011b; Wolfe, 1994). Several well-known findings have emerged from this work. Originally noted with symbol matrices, two important observations are that the target's absence is slower to detect than its presence (Chun and Wolfe, 1996) and that visual search time generally increases with "set size", the number of discrete elements in the stimulus (Palmer, 1995). The way in which the target is shown to the participant before the visual search trial is important also: image cues are better at guiding search than textual ones (Wolfe et al., 2004). Additionally, previous work considering finding elements of a specified shape, size, color, and quantity suggests that if the target is described in textual form, stating its color leads to better visual search performance, followed by its size and shape (Williams, 1967).

Notably, the visual search literature reports on a wide range of studies of the mechanisms by which targets guide search. Classic theories of visual search such as the Feature Integration Theory (Treisman and Gelade, 1980) and Guided Search (Wolfe, 2021) posit that, because of humans' limited capacity for object recognition, items in a visual field need to be selectively chosen for processing to identify whether they are the target. In these environments, feature guidance is an important mechanism, with a certain set of features, like color, motion, orientation, and size, efficiently guiding search; a review is provided by Wolfe and Horowitz (2004). Whereas guidance via top-down and bottom-up attributes prevails in studies with laboratory stimuli, natural scenes, containing objects that may exhibit hierarchical structure, benefit from additional search mechanisms (Henderson, 2005; Henderson and Ferreira, 2004). In one view, scenes have particular semantic meanings and syntax, just as grammar does (Henderson and Ferreira, 2004). This structure can be exploited for solving the search problem, for instance, if the target is a lamp, looking at tables is beneficial. Accordingly, natural scenes benefit from scene guidance (Biederman et al., 1973), wherein fixation locations are selected without regard for whether targeted features are visible (Wolfe, 2020). People tend to look at meaningful regions of images, like humans (Buswell, 1935; Yarbus, 1967), and evidence of relatively inefficient search when the target items are placed in idiosyncratic locations (Biederman et al., 1982) implies that scene guidance contributes to search of natural scenes.

2.2. Visual search in GUIs

Research conducted with GUIs has yielded findings similar to those with symbol matrices. One example is the growth of search time with absence of the target and the number of items displayed (Grahame et al., 2004; Trapp and Wienrich, 2018). In GUI settings, as in natural scenes, analyses pertaining to item count hinge on the definition of "set size", which may not be obvious. Therefore, several clutter metrics have been developed in place of count to describe crowdedness of visual displays (Rosenholtz et al., 2007; Miniukovich and De Angeli, 2015). A group of other metrics connected with visual complexity has been suggested to aid further in assessing usability of interfaces (Miniukovich and De Angeli, 2015). Research with GUIs has also produced evidence of textual elements being more difficult to find than images (Yuan and Li, 2020). This has practical implications, as a parallel can be drawn between looking for simple visual features and icon search: seeking of these GUI elements is often based on recalling basic visual features from memory (Kieras and Hornof, 2014). Furthermore, characteristics of the target aid in search, as varied colors and rectangular borders in the context of app icons illustrate (Liu et al., 2021), alongside differences from surrounding elements (Trapp and Wienrich, 2018).

Contributions in modeling visual search for predictive purposes is prevalent in HCI. Many of these models use observations from other contexts (Kieras and Hornof, 2014; Jokinen et al., 2020; Byrne, 1993; Halverson and Hornof, 2011). It is noteworthy too that, meanwhile, the field's empirical work on visual search is largely focused on specific GUI types or scenarios, such as menus (Bailly et al., 2014), icons (Liu et al., 2021; Byrne, 1993), keyboards (Jokinen et al., 2017), grids (Pfeuffer and Li, 2018), lists/groups (Halverson and Hornof, 2004; Hornof, 2004), webpages (Jokinen et al., 2020; Teo et al., 2012), and the aforementioned icons (Kieras and Hornof, 2014; Liu et al., 2021; Trapp and Wienrich, 2018). Among recent efforts sharing our aims are a large-scale study of visual search performance with GUIs (Yuan and Li, 2020) for the purposes of predicting visual search time using deep learning. However, that work focused on search times collected via crowdsourcing, without corresponding eye-tracking data; it was limited to webpages; and all trials had the target present. In contrast, several open-availability datasets, spanning multiple GUI types, exist, but these mostly come from free-viewing tasks, not visual search (Jiang et al., 2023; Kümmerer et al., 2024), see Table 1 for reference.

2.3. Selective attention and GUIs

Tasks wherein participants are allowed to look at GUIs without a specific goal, that is, engage in free-viewing, are often used to probe which regions of an image catch the eye. In HCI, free-viewing has been studied especially in the context of webpages (Shen and Zhao, 2014; Jiang et al., 2023), mobile devices (Leiva et al., 2020; Jiang

Table 1
GUI-related visual search and free-viewing datasets.

Reference	Task	Environment	GUIs	Participants	Trials
VSGUI10K (ours)	Visual search	Mobile UI Desktop UI Webpage	900	84	10,282
UEyes Jiang et al. (2023)	Free-viewing	Mobile UI Desktop UI Webpage Poster	1980	62	20,088
Leiva et al. (2020)	Free-viewing	Mobile UI	193	30	~4600
Shen and Zhao (2014)	Free-viewing	Website	149	11	N/A
Bailly et al. (2014)	Visual search	Linear menu	9	22	39,564
Jokinen et al. (2020)	Visual search	Consumer interface OS interface Website	3	20	24,514
Yuan and Li (2020)	Visual search	Website	N/A	1887	28,581

Table 2
The variables examined for correlations with search times. Controlled variables (indicated with *) were adjusted in the laboratory study, while the rest were allowed to vary per the naturalistic sample's composition. References to color variability (CV), clutter (CL), and grid quality (G) follow the practice of Miniukovich and De Angeli (2015). The "levels" are those present in the VSGUI10K dataset; naturalistic variables were computed via the Aalto Interface Metrics server (Oulasvirta et al., 2018).

Variable	Levels in data
Absence of the target*	Present
	Absent
Target cue*	Image
	Text + color
	Text
GUI category*	Mobile UI
	Desktop UI
	Webpage
PNG file size (CV1)	16 kB–2 MB
Distinct RGB values (CV2)	177–47,285
Distinct RGB values per dynamic cluster (CV3)	6.0–35.82
Static clusters (CV4)	32–16,978
Dynamic clusters (CV5)	1–2275
Contour density (CL1)	0.0–0.12
Subband entropy (CL2)	0.67–4.68
Feature congestion (CL3)	1.42–9.75
JEPG file size (CL4)	14 kB–436 kB
Number of visual GUI blocks (G1)	1–127
Number of alignment points (G2)	4–178
Number of block sizes (G3)	1–40
GUI coverage (G4)	0.0–0.71
Number of vertical block sizes (G5)	1–23
Figure-ground contrast	0.3–0.97
Contour congestion	0.23–0.9

et al., 2023), posters (O'Donovan et al., 2014; Jiang et al., 2023) and information visualizations (Borkin et al., 2016). Among the patterns found in free-viewing are a persistent upper-left bias in eye movements (Jiang et al., 2023; Leiva et al., 2020) and that people tend to look at text and images (Leiva et al., 2020). In addition to empirical work, computational modeling of saliency has been popular to predict which areas of a stimulus draw attention (Itti et al., 1998; Kümmerer et al., 2015; Cornia et al., 2018; Fosco et al., 2020). Distribution of attention in the context of GUIs has received some attention also in studies of browsing, where the tasks range from "open-ended" ones without specific goals to directed search. These studies have identified users' tendency to explore search-engine result pages (SERPs) and other pages in patterns that resemble the letter "F" (Nielsen, 2006; Pernice, 2019; Buscher et al., 2009) or a triangle (Hotchkiss et al., 2005; Hotchkiss, 2014). In addition to producing such well-known findings, previous work has focused on examining how users allocate attention to specific elements on the screen — for instance, their tendency of ignoring advertisement banners (Resnick and Albert, 2014, 2016). Finally, previous work offers design guidelines based on perception of elements on the display, for instance, tendency to use and seek structure or biases based on expectations and goals (Johnson, 2014).

2.4. Research questions

We examine the structure and special characteristics of visual search in GUIs through the following research questions:

1. How do relevant factors from previous work in symbol matrices (absence of the target, target cue, set size) and GUIs (GUI category, visual complexity) impact search times? Table 2 outlines our main independent variables drawn from these categories.
2. What similarities and differences arise in eye-tracking data of free-viewing and visual search tasks?

3. Method

We conducted an eye-tracking study to mimic various everyday visual search tasks in GUI settings. As the design space for these tasks is large, we make the following design choices to produce a sample that closely resembles the environments for which we set out to apply the results: We randomly sampled GUIs in three categories (webpages, desktop UIs, and mobile UIs) from the UEyes dataset (Jiang et al., 2023), representing a wide range of everyday-use interface screenshots collected manually or from existing datasets such as Visual Complexity and Aesthetics (Miniukovich and Marchese, 2020) and Rico (Deka et al., 2017). Also, we varied whether the target is present or absent and the way in which it is presented to the participant (in image or textual form). The study itself was conducted in a laboratory setting in order to collect high-quality eye-tracking data and to avoid nuisance factors. Each target was shown to the participant in a visual format, in line with conventions in other visual search experiments (Jokinen et al., 2020; Yuan and Li, 2020).

The collected VSGUI10K dataset, with eye-tracking data for 10,282 visual search tasks from 900 GUIs, represents a broad variety of stimuli in comparison to the publicly available datasets prepared with similar aims (see Table 1 for comparison). Attesting to the breadth of our dataset, Jokinen et al. (2020) considered only three GUIs, with 24,514 trials, while Yuan and Li (2020) covered 28,581 trials without any eye-tracking data and Brumby and Zhuang (2015) worked from 39,564 trials with just menus. We make the dataset available in conjunction with our experiment's pre-registration.[1]

3.1. Participants

In total, 84 people were recruited, through mailing lists, social media, word of mouth, and on-campus advertisements. The participants described themselves as follows:

- **Gender:** 40 self-identified as women, 42 as men, and three as non-binary. Multiple options were allowed.
- **Age:** 71 were aged 18–30, nine 31–50, and four over 50.
- **Occupation:** 51 reported that full-time student described them best, with the next most common statuses being full-time employment (23) and working part-time (21). Two participants had no paid employment, three marked that they were part-time students, five noted some other occupational status and one preferred not to answer. Multiple options were allowed.
- **Computer usage:** All participants reported using desktop (or laptop) computers and mobile devices at least weekly.
- **Vision:** The participants had normal or corrected-to-normal vision, and 59 reported wearing glasses/contact lenses or using other vision-correction.

Participants received 15€ in compensation for the experiment, which lasted approximately 60 min. The study was conducted in accordance with local procedures for ethics approval.

3.2. Materials

We obtained the stimuli from the UEyes dataset (Jiang et al., 2023), which contains screenshots from desktop interfaces, mobile ones, webpages, and posters. The difference between the UEyes data and ours is that the former were collected in free-viewing tasks wherein the user was instructed only to look at GUIs, instead of searching for specific elements. In addition, we excluded posters and selected 300 screenshots each from the remaining three categories for 900 GUIs in total, to

[1] Link to VSGUI10K dataset and pre-registration: https://osf.io/hmg9b/. The dataset is described in supplementary materials.

gather a broad sample for which data collection would still be feasible. Targets for the visual search tasks were selected by hand with Label Studio (Tkachenko et al., 2020–2022), with 1–6 targets being extracted for each image. Each target was chosen on the basis of pre-specified criteria, which only a few targets might fulfill in some cases. The targets, rectangular unique elements on the screen, each belonged to one of the following categories: input field, icon, menu item, widget, button, image, and title/subtitle. Most targets contained text.

3.3. *The task*

A visual search task comprised four displays (as Fig. 2 shows):

1. **Target description**: The target is displayed in the middle of the screen. A participant indicates preparedness to proceed with the task by pressing the spacebar. The description takes the form of one of three “levels”, denoted as Image, Text+Color, or Text.
2. **Fixation cross**: A black cross appears in the middle of the screen. The participant fixates on the cross for 1 s, to standardize where each of the visual search trials starts.
3. **Visual search**: The GUI screenshot is shown, and the participant performs the visual search task. Pressing the spacebar signals search completion. The maximum time for each trial is 30 s.
4. **Validation**: The outline of the GUI screenshot is displayed. In the lower-right corner of the screen, a rectangle with the text “Target absent” is shown. The participant is instructed to look at the location where the target was detected or, in its absence, look toward the lower-right corner. The participant is asked to gaze at the location for 3 s.

3.4. *Experiment design*

Our study followed a 3 × 3 × 2 within-subjects design. Participants were exposed to different levels of the three controlled variables: (1) GUI type (Webpage, Mobile UI, and Desktop UI), (2) target description (Image, Text, and Text+Color), and (3) absence/presence of the target. The GUIs were sampled from everyday GUIs in the UEyes dataset (Jiang et al., 2023). Additionally, our analysis below considers a set of independent variables related to the characteristics of the GUI — namely, metrics for visual complexity (see Table 2). The GUI images were organized into blocks of 18 visual search tasks each, which a pilot study demonstrated to yield a reasonable duration for each block. The blocks contained a mix of screenshots, from multiple GUI categories. A participant completed as many blocks as possible in the time allocated for the experiment (60 min). On average, participants were able to complete approximately seven blocks in that time. Each image received on average 11.42 trials.

The visual search tasks for the entire experiment were selected such that two targets were sampled for each screenshot (if available), and six instances of each target/screenshot pair were included in the overall sample. We set the target to be absent in approximately 10% of the trials. A Text or Text+Color description was provided in 50% of the trials (25% of samples in the full dataset specified text for the target, and a further 25% included a description of the background color in addition). The cue in the rest of the trials was an image. The visual search tasks’ order was randomized prior to collection of data. We made sure that the images were repeated at a low frequency, to avoid learning effects. In total, our dataset contains 10,282 visual search tasks (Figure 1 in supplementary materials give details on the composition of the trials). These visual search tasks were divided into blocks that contained 18 images each, preserving the randomized order. The code used in generating the stimuli and randomizing their order is publicly available alongside the dataset.

3.5. *Apparatus*

We conducted the experiment with a Gazepoint GP3 HD eye-tracker with a 150 Hz sampling rate and no head mount (Gazepoint, 2021). The GUI screenshots were shown on a monitor with 1920 × 1200 px resolution, at 94.3 ppi (HP Z24n 24-inch). Participants were instructed to sit 50–65 cm from the monitor. We used the integrated Gazepoint Control software for calibrating the eye-tracker, while Gazepoint Analysis was used to show stimuli and record data. The facilitator monitored the eye-tracking data for the duration of the experiment, and the eye-tracker was re-calibrated if drift was observed. The GUI screenshots were shown such that webpage ones spanned 75% of screen height (approximately 24°). Mobile UI stimuli were scaled to appear the same size as they would from their natural viewing distance corresponding to a visual angle of 19° or about 61% of screen height. Desktop UI ones were scaled to span 75% of the screen height at maximum.

3.6. *Procedure*

Before the experiment started, the participants were familiarized with it and signed an informed-consent form. They were allowed to adjust the positioning of the monitor and table to match their customary desktop usage patterns. Also, participants were able to adjust their distance from the screen (within the 50–65 cm range). These adjustments addressed the calibration that the eye-tracker required for guaranteeing high-quality data. The experiment began with the eye-tracker’s nine-point calibration via Gazepoint Control and testing by means of the calibration test screen. The calibration process iterated until at least eight accurately registered calibration points were acquired, and additionally the eye-tracker was calibrated mid-experiment if drift was observed. After calibration, the participant received a practice

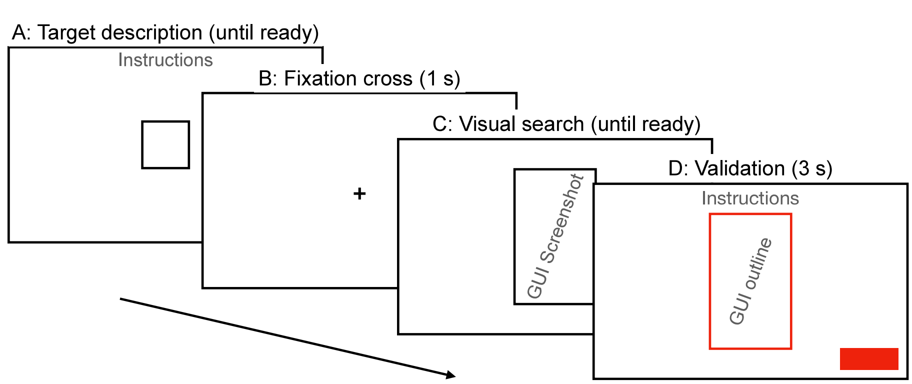


Fig. 2. A visual search trial. The target is the first thing shown to the participant (pane A), followed by a fixation cross to set the starting point of each search (pane B). Search is performed for a given GUI screenshot (pane C), and a separate screen is used to indicate where the participant detected the target via fixation at that location (pane D).

block of task. The practice block was designed to ensure participants' understanding of the experiment procedures and task requirements.

Each participant was allotted a one-hour session to complete as many blocks as possible, with self-managed breaks between the blocks. In every task (comprising the four screens shown in Fig. 2), the participant was asked to fixate on the centrally positioned cross in the second step for a span of 1–2 s and, in the fourth step, on the location of the target found for 3–4 s, for making sure of the calibration's accuracy and verifying the result of localizing the target. Thereby, we separated validation of whether the target was found from the search. This method of validating where the participant had detected the target was chosen to avoid confounding the visual search results with the reaction times that clicking involves and due to limitations of the eye-tracking software. A metronome was active throughout the experiment to aid participants in gazing the screen for the specified durations. Each participant answered questions related to demographic details after the last block of GUI images was shown.

3.7. Data analysis

Next, we briefly describe how we processed the data, segmented the GUI screenshots (for set-size), and estimated coverage used in analyses.

Data processing. We used the fixation data provided by Gazepoint's custom fixation filtering (FPOGX and FPOGY, for "fixation point of gaze") in further analyses. All analysis used (only) the fixation data from the visual search screen (pane C in Fig. 2), except where we indicate otherwise; fixation data from the validation screen constitute an exception, in light of the need to examine the number of successful trials. All fixations were deemed valid by Gazepoint's validity metric (FPOGV). Upon collection of all the data, fixation points were detected as missing from the visual search screen for 80 images, and we ran these trials with new participants to avoid learning effects. Total search time was calculated in two ways: (1) as the trial's duration and (2) as the sum of fixation durations (via FPOGD). Our reporting uses the results from metric 1 (the interpretation of the results is similar for sum-total fixation duration). In total, 61 trials exceeded 30 s, but were not removed from the dataset. We de-biased the eye-tracking data through the logs from the fixation-cross step (pane B in Fig. 2), where these were available for the trial, by subtracting the error along the x- and y-axis from the fixation coordinates. While we examined errors made during the searches (i.e., mistakes in localizing the target), we have not removed any trials from the dataset (see Figure 12 in supplementary materials). Approximately 90% of the trials where within a minimum normalized distance of 0.2 from the target in the validation step. There were 51 duplicate trials, for which the second instance was removed. In total, the dataset encompasses 10,282 visual search trials.

Segmentation. Elements within each GUI were segmented as described by Jiang et al. (2023). When segmenting the interfaces, the elements were classed into three categories: image, text, and face. To accomplish this, an extended UIED model (Xie et al., 2020) was used, initially designed for the identification of images and text within GUIs. For calculating set sizes, we counted all image and text elements within a GUI. Examples of segmentations are provided alongside the dataset.

Foveal area. We estimated the area of the participant's foveation on the GUI assuming that they sat 50 cm from the screen and that the monitor operated at 94 ppi. Thus, the foveal area's radius was assumed to be around 32 px. We then approximated overlapping foveal areas, using the Python `Shapely` package (Gillies et al., 2007–2024).

3.8. The multilevel models

We used multilevel models to investigate which variables explain search times. Three such models were fitted to consider variables that (1) were controlled in the experiment design, (2) related to visual complexity, and (3) were significant in models 1–2 and hypothesized to interact with absence of the target (also, the supplementary materials present models constructed for alternative independent and dependent variables, see Figures 13–20). Multilevel models, or mixed-effects models, were chosen since they can factor in the data's hierarchical relations, which involve measurements of individual search times (level 1) nested across the various stimuli (level 2) and further nested for distinct participants (level 3). Our multilevel modeling included random effects for stimulus and participant ID, accounting for the variability in search time that may be attributed to these higher-level groupings.

For model fitting, the `lmer` function from R's `lme4` package (Bates et al., 2015) was used. Hypothesis tests for the fixed effects in our model were conducted via the Satterthwaite degrees-of-freedom method as implemented in the `lmerTest` package (Kuznetsova et al., 2017). The `sjPlot` package was used in visualizations and analysis (Lüdecke, 2023). We also provide Standardized Beta Coefficients, serving for estimation of effect sizes, using the following thresholds: $\beta \leq 0.29$: small, $0.29 < \beta \leq 0.49$: medium, and $\beta > 0.49$: large (Cohen, 1988; Nieminen, 2022). Our use of multilevel models considers the within-group multicollinearity via the random part of the model. In examining effects, we employed comparisons of estimated marginal means, which, analogously to ANOVA, is not strictly considered *post hoc* testing. The modeling code has been released in conjunction with this publication. The individual models are described in turn below.

The controlled-variable model. Since our experiment followed a $3 \times 3 \times 2$ design, the controlled-variable model takes three independent variables: category, cue, and absence.

The visual-complexity model. To evaluate the impact of individual screenshot characteristics on search times, we used automated GUI evaluation metrics, computing them by using the Aalto Interface Metrics (AIM) server (Oulasvirta et al., 2018). We took an approach similar to Miniukovich and De Angeli (2015), wherein the metrics are categorized into those pertaining to (1) visual clutter, (2) color variability, (3) contour congestion, (4) figure–ground contrast, and (5) layout quality. Of these classes, metrics for visual clutter (CL), color variability (CV), and layout quality (G) are multi-item ones, which were combined via factor analysis following the prior method (Miniukovich and De Angeli, 2015). Specifically, the `factanal` function of R was used for factor analysis and computation of factor scores, with number of factors evaluated using the `n_factors` function from the `parameters` library (Lüdecke et al., 2020).

The interaction model. Proceeding from the models related to controlled variables and visual complexity, we identified those coefficients and corresponding independent variables that had a significant p-value. To describe interactions with absence of the target, we constructed a third model, with contributions from analyses that considered factors 1, 3, and 4, figure-ground contrast, contour congestion, category and cue as independent variables, where each was tested against the assumption that it interacted with absence of the target. We also considered a version of the interaction model that takes into account set size instead of the visual complexity metrics (see "Segmentation" for its computation), to render comparisons with previous results involving symbol matrices and natural scenes easier (see supplementary materials).

4. Results

4.1. Fitted linear mixed-effects models

As is detailed in Section 3.8, our analysis builds on three separate models, considering different independent variables, associated with (1) controlled variables in the experiment design, (2) visual complexity, and (3) interactions with absence of the target. This division functions to simplify the interpretation of our models.

Fig. 3. Fixed effects for the controlled-variable model when search time is the dependent variable. Random effects are considered for each image and participant ID. The figure provides absolute (pane A) and standardized (pane B) estimates both. Significance levels for the coefficients are indicated thus: $*** = p < 0.001$, $** = p < 0.01$, $* = p < 0.05$. Webpage stimuli serve as the baseline for GUI category, and images are the reference among cues. Increases in search time with respect to the baseline are indicated in red and decreases in blue.

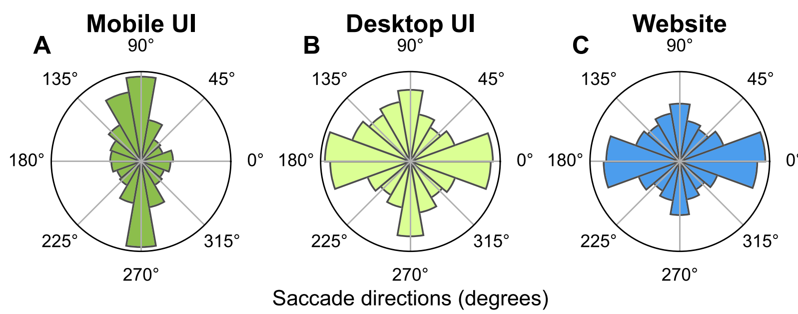


Fig. 4. Saccade directions during search, by GUI category. Scanning of GUIs tends to follow their longitudinal axis.

The controlled-variable model. Absence of the target, as well as presenting its cue in textual format over an image increase search times. Also, search is faster in desktop and mobile UIs than in webpages (see Fig. 3 for fitted coefficients). Specifically, search time increased by 6.97 s ($SD = 0.15$) with absence of any target. This is the largest effect in the data (1.45 with $SD = 0.03$). Providing descriptions of the target in purely textual form led to more protracted searching than image-based ones — search time rose by 1.45 s ($SD = 0.09$) and 1.41 s ($SD = 0.09$) for the Text and Text+Color condition, respectively. These coefficients indicate medium effect sizes (0.30 and 0.29 both with $SD = 0.02$ for Text and Text+Color, respectively). Relative to webpages, searches with desktop and mobile interfaces, respectively, were 0.78 s ($SD = 0.16$) and 2.23 s ($SD = 0.16$) faster, showing small to medium-sized effects (−0.16 with $SD = 0.03$ for the former, −0.47 with $SD = 0.03$ for the latter). Partly, this effect can be attributed to the interfaces' size; in general, webpages have the largest search area, followed by desktop and then mobile UIs (see Figure 2 in supplementary materials). The marginal and conditional R^2 values for this model are 0.235 and 0.420.

Target absence exerts a significant influence on search times. The longer searches accompanying absence can be attributed in part to covering the GUIs more exhaustively in trials where the target is absent (i.e., to a larger estimated total foveation area; see Fig. 5). That is, users tend to foveate on more of the GUI when the target is absent. Additionally, coverage is higher for mobile UIs than the other two categories; given the faster search with mobile GUIs, this is probably linked to the size of the space: mobile UIs are smaller and hence support quicker scanning. Additionally, saccade directions for mobile UIs exhibit a vertical-scanning pattern, suggesting that mobile UIs can be relatively efficiently searched by working one's way along the length dimension. However, search in desktop UIs and webpages shows a tendency to scan widthwise. That is, with website and desktop UIs, which typically have a landscape orientation, scanning occurs via horizontal eye movements (see Fig. 4, panes B and C).

The visual-complexity model. For the independent variables of the visual complexity model, two separate factor analyses were conducted to determine loadings, one for visual clutter and color variability and the other for layout quality. Loadings resulting from the factor analysis are characterized in . Similar to work by Miniukovich and De Angeli (2015), the first analysis revealed three factors (labeled 1–3 in the table). One of these has higher loadings for visual-clutter metrics (Factor 1 in), the other two for color variability. In the factor analysis focusing on layout quality, one factor emerged. On account of this analysis, we proceeded to use factors 1–4, plus contour congestion and figure–ground contrast, when fitting the visual-complexity model, and compute factor scores for each.

Table 3
Factor loadings resulting from the factor analysis performed for visual complexity metrics. Following the literature's lead (Miniukovich and De Angeli, 2015), we evaluated the metrics in two factor analyses: for (1) color variability (CL) and visual clutter (VL) and (2) grid quality (G). Factor loadings higher than 0.7 are gray-highlighted.

Loading / Metric	Factor 1	Factor 2	Factor 3	Factor 4
PNG file size (CV1)	0.332	0.239	**0.733**	–
Distinct RGB values (CV2)	0.060	**0.908**	0.387	–
Distinct RGB values per dynamic cluster (CV3)	0.039	0.234	**0.804**	–
Static clusters (CV4)	0.398	0.565	0.144	–
Dynamic clusters (CV5)	0.041	**0.970**	0.228	–
Contour density (CL1)	**0.884**	0.123	0.206	–
Subband entropy (CL2)	**0.872**	0.249	0.248	–
Feature congestion (CL3)	**0.985**	0.005	0.017	–
JEPG file size (CL4)	0.498	0.486	0.051	–
Number of visual GUI blocks (G1)	–	–	–	**0.890**
Number of alignment points (G2)	–	–	–	**0.998**
Number of block sizes (G3)	–	–	–	**0.916**
GUI coverage (G4)	–	–	–	0.194
Number of vertical block sizes (G5)	–	–	–	**0.781**

In comparison to the controlled-variable model, the visual-complexity model captures generally smaller effects (see Fig. 7). Contour congestion and factors 1 (visual clutter), 3 (color variability), and 4 (grid quality) contribute significantly in a statistical sense to

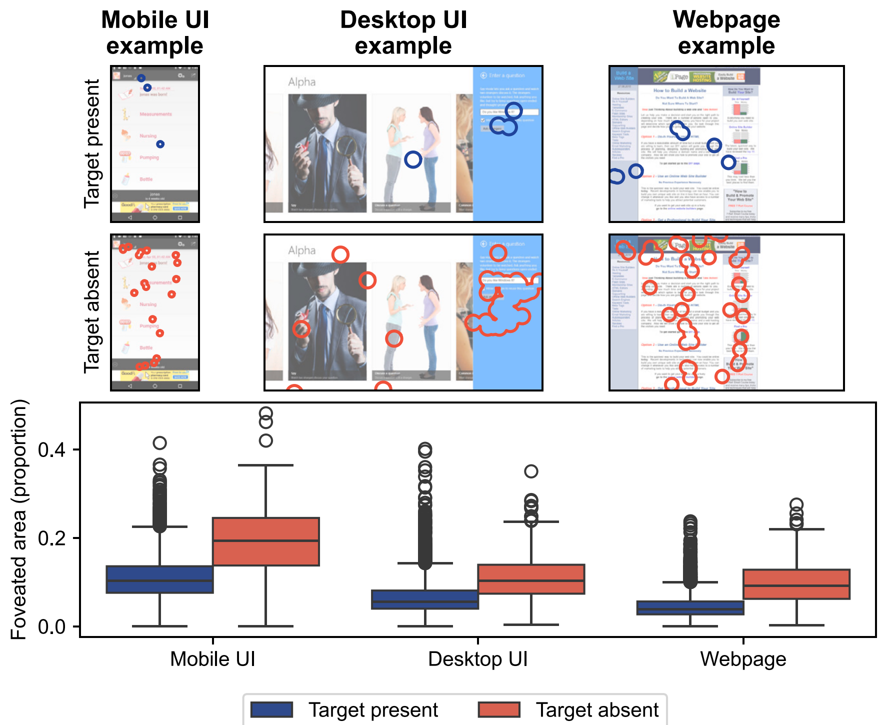


Fig. 5. Coverage of the foveated area in searches. Coverage was greater with mobile UIs and when the target was absent. GUI examples with fixations from the VSGUI10K corpus are shown in the top two rows assuming viewing distance of 50 cm.

Fig. 6. Examples of GUIs from VSGUI10K corresponding to specific metric or factor-score values. Also shown is the given metric's distribution in the GUI corpus.

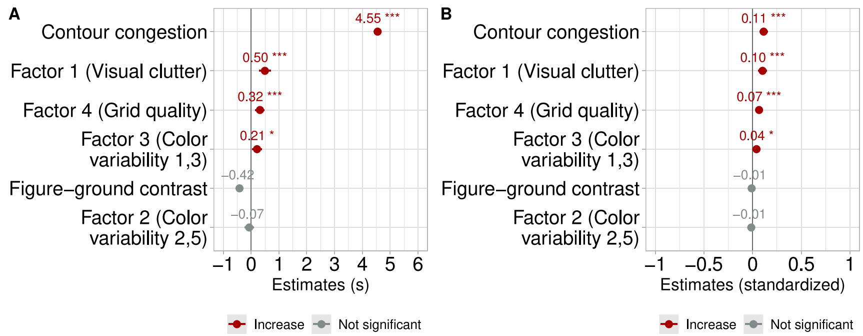


Fig. 7. Fixed effects for the visual-complexity model when search time is the dependent variable. Random effects are considered for each image and participant ID. The two panes provide absolute and standardized estimates both. Factor names reflect the highest loadings. The significance levels for the coefficients are denoted thus: $*** = p < 0.001$, $** = p < 0.01$, $* = p < 0.05$. Increases in search time are indicated in red and non-significant coefficients in gray.

explaining search times, while figure–ground contrast and Factor 2 (another color variability metric) do not. Search time rises with contour congestion, though the effect is small (0.11 with $SD = 0.02$). Factor 1's highest loadings (>0.7) are on contour density, subband entropy, and feature congestion, all related to interface clutter. Therefore, the fitted coefficient for Factor 1 suggests that search time grows as clutter does, with a small effect size. With Factor 3, in contrast, the highest loading (>0.7) is on PNG file size and number of distinct RGB values per dynamic cluster, suggesting that search time increases with these figures. That said, the effect is very small (0.04 with $SD = 0.02$). Finally, the fitted coefficient for Factor 4, connected with grid quality, points to a small effect (0.07 with $SD = 0.02$) in which searches take longer as the grid gets more complex. The values for R^2 imply poorer fit for the visual-complexity model than the controlled-variable one, with a 0.053 marginal and a 0.308 conditional R^2 value.

The interaction model. Absence of the assigned target has been shown to affect search times significantly (Chun and Wolfe, 1996). That observation is echoed in the fitted controlled-variable model (see Fig. 3). Therefore, we investigated any interactions between the target's absence and the other variables. Fig. 9 shows the fitted coefficients for the interaction model. The figure excludes those variables that were not significant in the controlled-variable and visual-complexity models. Note also that the categorical variables in our analysis take the following reference levels: Webpage for the GUI-category variable and Images for cues. The fitted coefficients of the interaction model largely follow the trends visible under the controlled-variable and visual-complexity models. Recognizing the complexity of interpreting interaction effects within a multilevel model, we illustrate the results via the overview in Fig. 8. Results are plotted for the coefficients from the interaction model, with webpage stimuli and Image cues as the references for categorical variables. To aid in comparing our results to previous findings, we also fitted an interaction model that takes not complexity metrics but set size as the independent variable (see Figure 17 in supplementary materials). Furthermore, that material provides a generalized linear mixed-effects model with gamma-family distribution (log link) that addresses the same variables alongside residual plots. The discussion below flags any deviations between the linear and generalized modeling.

In our data, the coefficients for mobile UIs, cue types, factors 1 and 4, and contour congestion remain significant when the target is present. Search times in the desktop-UI condition do not significantly differ from those for webpages when the target is present. In mobile UIs, search is faster by 1.00 s ($SD = 0.19$), with a small effect size (-0.21 with $SD = 0.04$). Presenting the cue as an image leads to better search times than presenting it in a text-featuring format when the target is present: search is 1.38 s slower ($SD = 0.09$) and 1.28 s slower ($SD = 0.09$) for the Text+Color and Text condition, respectively, in small to medium-sized effects (0.29 with $SD = 0.02$ and 0.27 with $SD = 0.02$). Factor 4, encompassing the grid-quality metrics, shows a small effect (0.05 with $SD = 0.01$) corresponding to a 0.25 s increase in search time ($SD = 0.06$) with higher-complexity grids. When the target is present, search time rises by 2.59 s ($SD = 0.68$) with a unit increase in contour congestion; the effect size here is small (0.06 with $SD = 0.02$). Factor 1, putting the highest loadings on visual-clutter metrics, likewise shows a small effect, wherein search times grow by 0.42 s ($SD = 0.08$). The effect size is 0.09 with $SD = 0.02$. Additionally, Factor 3 no longer offers a significant coefficient. To illustrate the units of these factors, which may be otherwise difficult to interpret, we refer to Fig. 6. Analysis with set sizes (Figure 17 supplementary materials) suggests that the effect of including more items is small when the target is present (0.06 with $SD = 0.02$), with a cost of 0.02 s per item.

Absence of the target again shows a large effect size (1.52 with $SD = 0.07$). In the linear model, some interactions with the target's absence prove significant. When the target is absent, search of desktop UIs is 1.03 s faster than webpage search ($SD = 0.40$), with a small effect size (-0.22 with $SD = 0.08$), while search of mobile ones is 1.83 s faster ($SD = 0.48$), with a medium effect size (-0.38 with $SD = 0.10$). The pattern for target cue looks different when the target is absent. With a Text cue, search is slower by 2.14 s ($SD = 0.32$), in a large effect (size 0.45 with $SD = 0.07$), while for a Text+Color cue there is no significant difference. Factor 4 too has an effect: when the target is absent, search time is projected to increase by 0.99 s ($SD = 0.17$), with a small to medium effect size (0.20 with $SD = 0.03$). Additionally, every unit increase in contour congestion yields a 4.91 s ($SD = 1.62$) increase in search times, although the effect is small (0.12 with $SD = 0.04$). Factor 1 shows a small effect (0.10 with $SD = 0.04$) when the target is absent, as search times increase by 0.48 s ($SD = 0.20$). As for set size, the cost of including another item when the target is absent is 0.04 s (see Figure 17 in supplementary materials). The effect size here is small (0.11 with $SD = 0.02$). It should be noted, however, that the interaction effects show a different pattern when the generalized model is applied, so one should interpret them with caution (see Figure 16 in supplementary materials). That is, while the main effect wherein the target's absence makes search slower persists, whether that absence exacerbates these effects requires further examination.

4.2. Comparison to free-viewing

Eye movements early in each trial led disproportionately toward the upper-left quadrant of the screen, with this bias being visible across all GUI types (see Fig. 10, panes A–D). The distribution of saccade directions (see Figure 26 in supplementary materials) supports positing the existence of such a tendency: the first fixations tended to move upward and leftward from the starting point at the screen's center. Our results suggest that the upper-left bias over the first few fixations is not sensitive to the target's location — irrespective of its position, the early fixations tend to be biased toward the upper left.

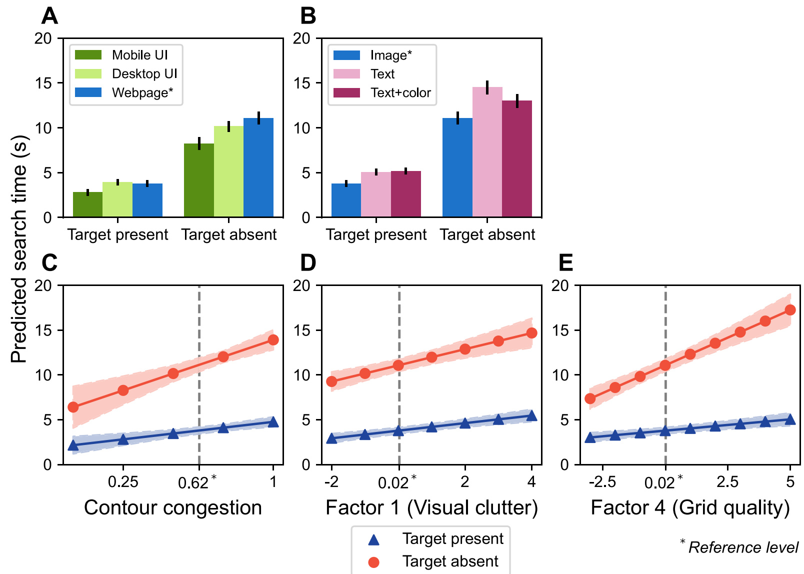


Fig. 8. Search times predicted by the fitted interaction model. Search was fastest in mobile UIs, followed by desktop and webpage interfaces. Image cues too produced faster searching. Also, contour congestion, clutter, and grid quality affected search times.

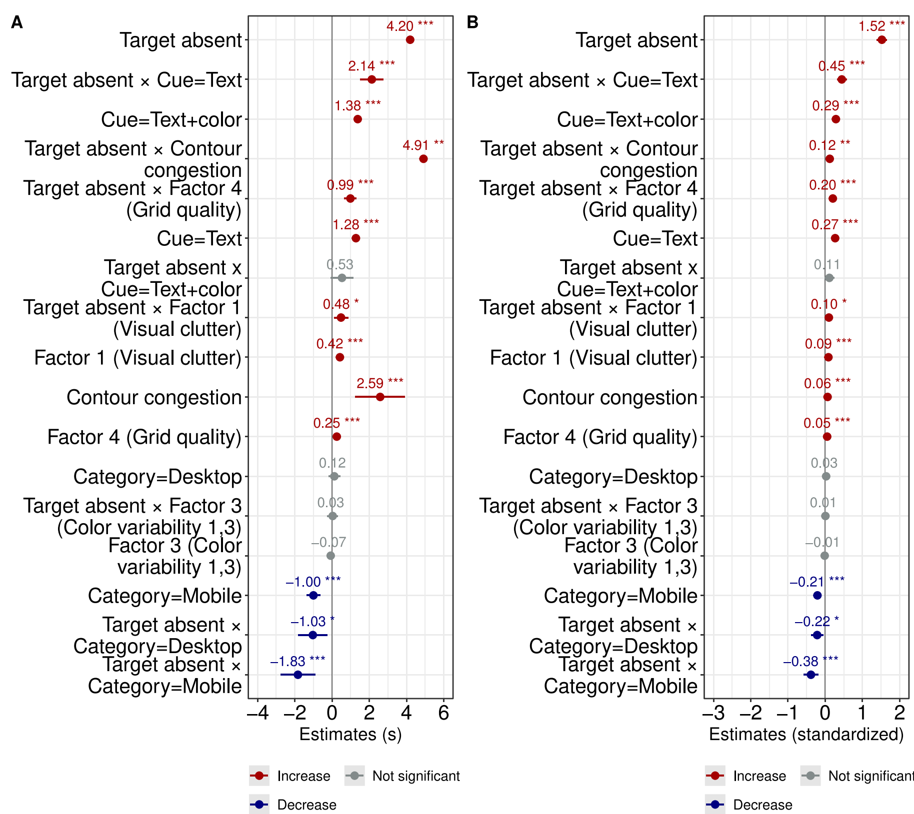


Fig. 9. Fixed effects for the interaction model when search time is the dependent variable. Random effects are considered for each image and participant ID. In the figure, showing absolute (pane A) and standardized (pane B) estimates, the factors are named in line with the highest loadings. Significance levels for the coefficients are indicated thus: ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$. Note that webpages serve as the baseline for GUI category, and images are the reference among cues for the categorical variables. Increases in search time with respect to the baseline are indicated in red, decreases in blue and non-significant coefficients in gray.

As expected, later fixations cluster more in the positions of the targets, their distribution, presented in Fig. 10, panes E–H. Stepping back to compare the distribution of fixations within VSGUI10K to that from the UEyes dataset (Jiang et al., 2023), we found upper-left bias visible across all categories (see Fig. 10, Q–T), with this bias persisting in the free-viewing task after the early fixations. While the concentration of the first fixations in our study's target-absent trials likewise points to an upper-left bias, this dissipated as participants' search progressed (see Fig. 10, panes M–P). The directionality of the saccades echoes this pattern: in this condition too, the first few tended to be directed toward the upper-left corner. Later search, in aggregate, covered the search area fairly evenly.

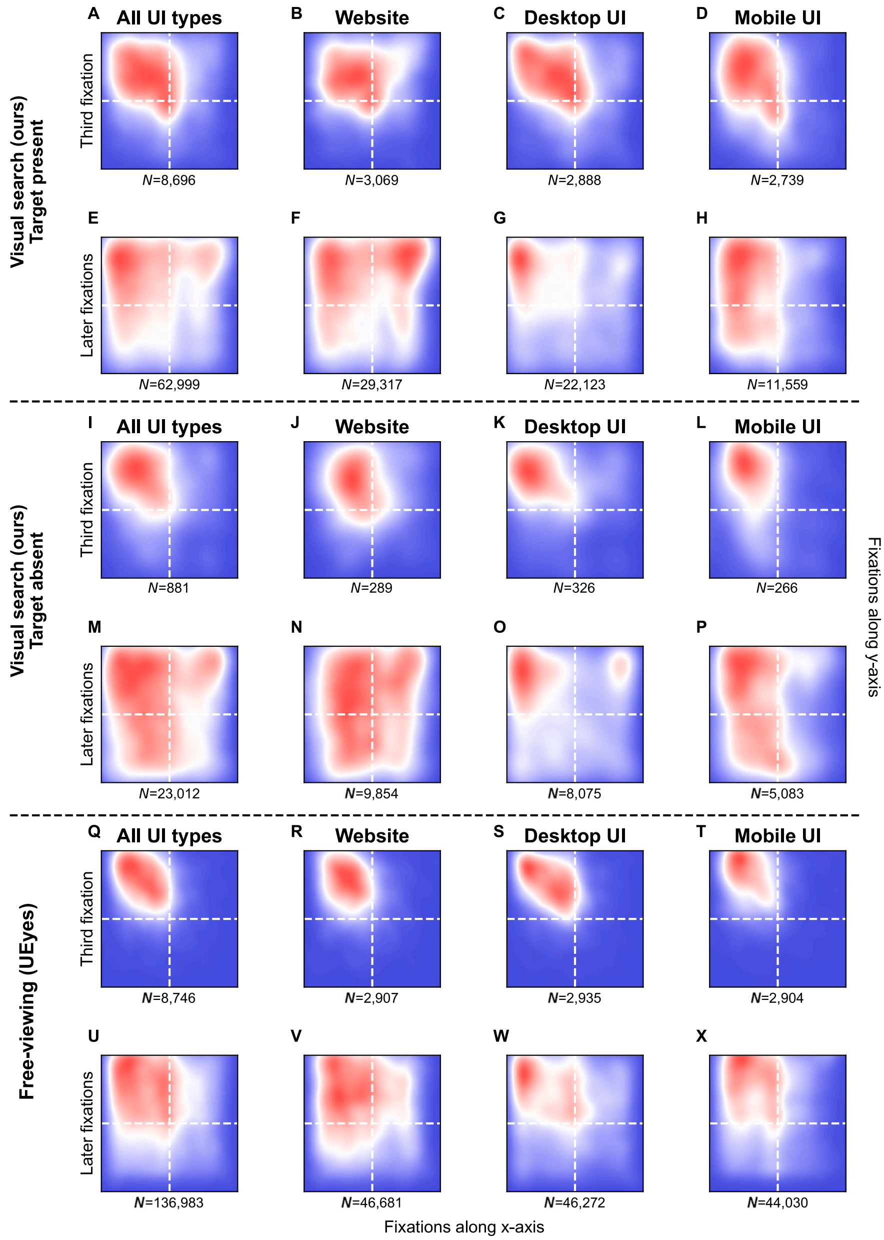


Fig. 10. Upper-left bias in distributions of fixations. An upper-left bias of fixations was present early in the trial, but this effect wanes later in the search. In the free-viewing setting examined by Jiang et al. (2023) the bias persisted.

5. Three stages of visual search

Contingent on our results, we posit that visual search in GUIs conforms to a three-stage pattern. In this parallel to a view from Just and Carpenter (1976), each stage has its own goal and principle of visual guidance:

1. **Guess**: The first few fixations take the gaze toward the top-left guided more by expectations than by the design.
2. **Scan**: Selective fixation on target candidates starts, guided jointly by the (GUI) scene's and target's features, and the gaze moves toward the target.
3. **Confirm**: The target is identified as correct.

5.1. Guess

The process begins with an opportunistic stage that is independent of the GUI's design: **Guess** (stage 1 in Fig. 11 and first deciles in Fig. 12) commences upon exposure to the GUI and accounts for the earliest fixations. Our results indicate that behavior in this stage shows a tendency to glance mainly at the upper-left quadrant of the screen. Studies with free-viewing and SERPs have revealed a similar bias (Leiva et al., 2020; Jiang et al., 2023; Nielsen, 2006; Hotchkiss et al., 2005; Hotchkiss, 2014), which is visible also in some laboratory experiments using symbols and natural objects as display items.

We surmise that this tendency is a result of limited information about the current GUI: upper-left is a good guess for the location of the target, or information about it, based on prior exposure to GUIs. Previous work provides parallels to this view, suggesting that this bias is either a part of a deliberate search strategy or a result of learned oculomotor patterns, like, reading from left to right (Chen and Zelinsky, 2006; Zelinsky, 1996). Leiva et al. (2020) link the bias to designers placing information that helps orient the user – for instance, logos, instructions, and headers – toward the display's upper left. Upon repeated exposure, users may form an expectation as to the location of this information. We would contend that the upper-left bias is likely

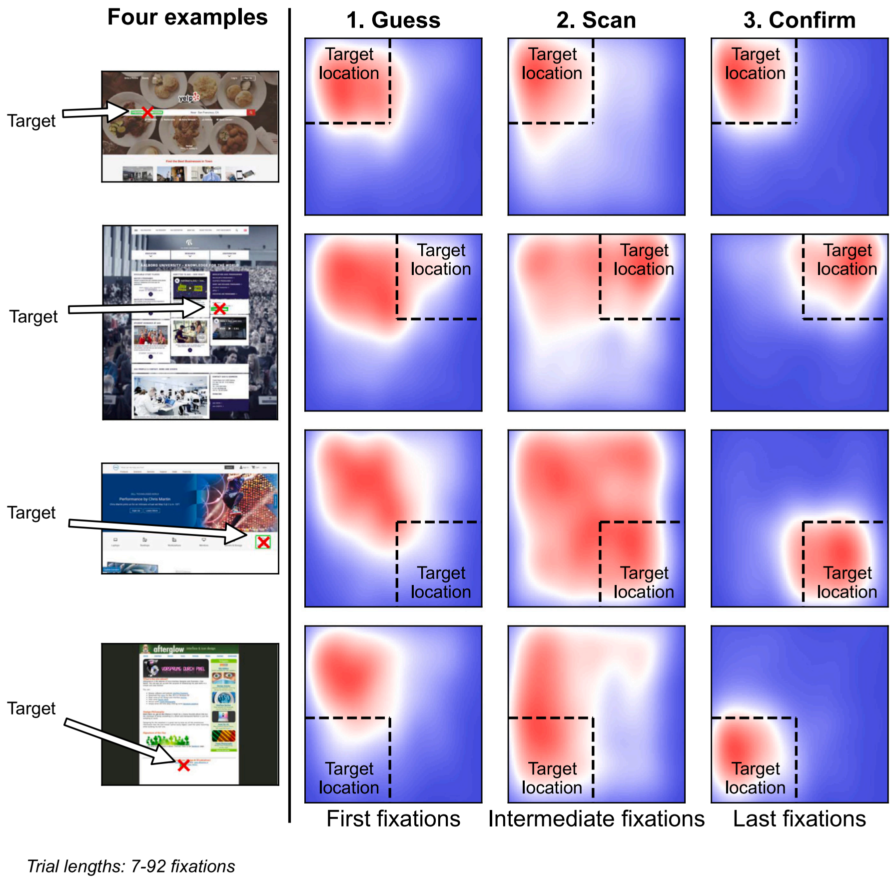


Fig. 11. Our results suggest that visual search in GUIs follows a pattern of Guess—Scan—Confirm, wherein directing the first fixations toward the upper left (1) is followed by a search of the GUI (2) and finally identification of the target location (3). The heatmaps here exclude the first two fixations (since trials always started at the center of the screen), and only trials with more than seven fixations are shown to include at least one fixation in each step (the supplementary material's visualizations include an alternative way of segmenting fixation paths). GUI and target examples from the `VSGUI10K` corpus also shown.

rooted in a combination of these influences: the user allocates initial attention to the top-left since this is both a good guess for the location of the target and an area suitable for gathering clues for navigating the GUI. Though users do not expect to find every item in the top-left (Roth et al., 2010), sampling the upper-left corner first is rational, because it allows quickly obtaining information from the region that is likely to contain, if not the target, at least many of the GUI's key elements. The term "bias" does not denote irrational behavior; rather, it reflects sensitivity to the statistical distribution of GUI elements. The pattern of bias is more pronounced in longer trials (as Fig. 12 shows). One explanation for this observation is that contextual cues are more necessary when searching is bound to take longer, while the solution in easier tasks is more obvious.

5.2. Scan

Once some information about the given GUI has been sampled, the second stage (stage 2 in Fig. 11 and middle deciles in Fig. 12) begins. In this stage, **Scan**, the user's gaze moves nearer the target item as candidate items get scanned. Our data show that, on average, modern GUIs are well-designed, in that their design guides the scanning stage well. While our results point to a tendency for increases in contour congestion, visual clutter, and grid complexity to slow the search, these effects are generally small. A separate analysis where set size was considered an independent variable led to a magnitude of growth in search times that was similar to those with the visual complexity metrics. With our GUI corpus, we observed that the cost of displaying a new element (adding a distractor) is just 20 ms (when the target is present). In comparison, with symbol matrices the effect of set size has been reported to range from near 0 ms/item in "feature" searches, wherein the characteristics of the target can contribute effectively to guiding the search, to 20–40 ms/item in "conjunction" searches, in which this is not possible (Wolfe et al., 2010, 2004). There is evidence that natural scenes support highly efficient searching, with search times increasing by a mere 3–10 ms/new item (Wolfe et al., 2011a). In contrast, when every candidate *must* be fixated upon, the increment nears 125–250 ms/item (Wolfe et al., 2010), approaching the average duration of one fixation. Evidence provided by Trapp and Wienrich (2018) supports this view, suggesting that the set-size effect's magnitude lies in the 6–75 ms/item range with GUIs. We also see evidence that users need only foveate on a relatively small portion of the GUI to solve their task. On the other hand, there is room to improve. Were display design perfect in the sense that targets are so easy to find that addition of items does not matter, the slope would be closer to zero (Wolfe et al., 2010); this has indeed been observed in some cases examined with symbol matrices, when certain features of the target guide search very well. Finally, we observed that image cues yield quicker searches than textual ones, a phenomenon consistent with observations with symbol matrices (Wolfe et al., 2004).

5.3. Confirm

In the last stage (stage 3 in Fig. 11 and last deciles in Fig. 12), users make sure their selection is correct. We denote this as the **Confirm** stage. In our data, the quadrant containing the target tends to attract more fixations quite late in searches. This implies that certain confirmation behavior takes place near the target location.

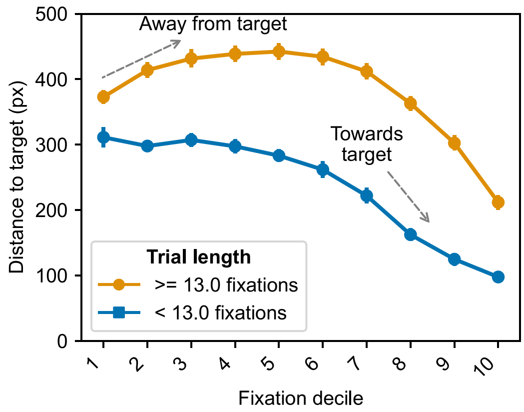


Fig. 12. Longer trials display a more pronounced Guess stage in search. Trials divided between those shorter than and at least 13 fixations (median). Similar plot shown for the axes and target locations separately in Figure 25 of supplementary materials.

6. Discussion

In 1994, Wolfe asked whether the rules of visual search articulated in studies with symbol matrices apply for natural scenes (Wolfe, 1994). Continuing in a similar vein, we have shed light on whether effects identified from symbol matrices and natural scenes apply in a wide range of everyday GUIs, by means of VSGUI10K – a dataset we collected for 10,000-plus visual search trials.

In summary, our results show overlap with those obtained in the other two environments: absence of the target is detrimental in terms of search time, and having an exact pictorial representation of the target makes search faster. However, the characteristics of the GUI as measured via visual-complexity metrics tend to have a fairly small impact on search times. We also noticed an upper-left bias in early search, similar to that present in free-viewing (Leiva et al., 2020; Jiang et al., 2023), but this effect fades over the search. The latter observation and our other results informed a synthesis of our findings into a three-stage Guess-Scan-Confirm search pattern. The first fixations directed to top-left serve as a *guess* of target location, or of the type of GUI, which aids in later search, wherein the following fixations perform a target-guided *scan* over the GUI. Finally, the last fixations *confirm* that the target was indeed located.

Absence of a target represents an interesting special case, with markedly higher search times. In fact, the target's absence accounts for the largest effect across all our data. This supports the absent-target–longer-search link reported in work with natural scenes and the limited prior studies with GUIs, wherein absence of the target led to search-time increases with a magnitude of several seconds (Neider and Zelinsky, 2008; Grahame et al., 2004). With symbol matrices too, absence of the target tends to produce higher search times in settings that demand, for example, serial deployment of attention or foveation for items; for instance, there might be no information available on any basic attribute, such as target color, that could guide efficient search (Wolfe et al., 2010). Our data suggest that the prolonged searching stemmed from the users extending their gaze over a larger area of the GUI. When the target is absent, later fixations generally spread across the entire GUI. Furthermore, every item added to the set in the target-absent state increased search times by 40 ms more than when the target was present. Absence of the target also seems to interact differently with the target's description type, although this effect is somewhat ambiguous and should be examined further in future work. Namely, a less pronounced difference in effect between image and text-plus-color hints when the target is absent suggests that the latter gives the user a better sense of when to abandon the futile search, while a mere sense of the target text seems to be a less useful signal.

Practitioners can use our linear mixed-effects models to quantify changes in search times in various kinds of use scenarios. Real-world GUI design entails several tradeoffs. For instance, our results reveal a tension in balancing between an overly complex GUI and omitting useful items: if a highly simplified GUI lacks items that users are looking for, this proves detrimental to visual search, while the cost of additional items seems low from our data. Managing these tradeoffs is generally left to designer intuition, yet a multi-level regression model is one possible tool for quantifiable insight for optimizing GUIs when multiple factors are at play. Our model allows assessing the effects of pertinent guidelines by testing multiple scenarios. With it, one could ask what happens if, say, the target is absent, the user is assumed to have a precise pictorial impression of the target, and the setting is a mobile GUI containing 35 items. We see potential for the VSGUI10K dataset as a useful resource for computational modeling of visual search, with particular value for exploiting machine-learning techniques that require sizeable datasets, not generally produced through controlled studies.

6.1. Limitations and future work

While we release a rich dataset for future research in this area, some limitations should be taken into account in analyzing that dataset and in interpreting our findings. Firstly, the participants in the controlled study were mostly young and students. Future work should verify that the results generalize to more representative populations, including more variety in the age distribution. Also, participants performed visual search tasks for an hour, which may have created some fatigue, and the metronome running in the background to help them gauge dwell times might have been distracting. Additionally, a more realistic task instead of pattern-matching (e.g., relying on memory) should be used in future work to understand any confounds. We also showed mobile GUIs on a desktop monitor instead of a phone; that might affect search behavior. Since the targets were sampled from everyday real-world GUIs, their locations display some bias that should be noted in future work using our dataset (see supplementary materials). The presented multilevel models have some limitations: while they provide informative quantifications of effects, their linearity renders their prediction power limited. Future work could apply and extend VSGUI10K to study non-linear and machine-learning-based approaches to modeling visual search. Finally, future work should delve into the three-stage pattern articulated and quantify it, in particular to understand whether the observed upper-left bias is a result of habits, prior exposure to GUIs, deliberate search strategy or something else.

CRediT authorship contribution statement

Aini Putkonen: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Yue Jiang:** Writing – original draft, Formal analysis. **Jingchun Zeng:** Writing – original draft, Investigation, Data curation. **Olli Tammilehto:** Writing – original draft, Investigation. **Jussi P.P. Jokinen:** Writing – original draft, Methodology, Formal analysis. **Antti Oulasvirta:** Writing – original draft, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work received support from the Research Council of Finland: the flagship program FCAI (grants 328400, 345604, 341763), Human Automata (grant 328813) and Subjective Functions (grant 357578). Department of Information and Communications Engineering at Aalto University also supported this work. AP Additionally received support from Aalto University School of Electrical Engineering. JJ was supported by Research Council of Finland (grant 330347). We thank our colleagues for feedback on the manuscript, particularly Luis Leiva, Robin Welsch and Francesco Chiossi for providing comments during early stages of the work, and Anna Shefl for proof-reading assistance.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijhcs.2025.103483.

Data availability

The dataset is available in OSF https://osf.io/hmg9b/, alongside the pre-registration of the study and a link to the code repository.

References

Bailly, G., Oulasvirta, A., Brumby, D., Howes, A., 2014. Model of visual search and selection time in linear menus. In: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, pp. 3865–3874. http://dx.doi.org/10.1145/2556288.2557093.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Biederman, I., Glass, A., Stacy, E., 1973. Searching for objects in real-world scenes. J. Exp. Psychol. 97 (1), 22–27.

Biederman, I., Mezzanotte, R., Rabinowitz, J., 1982. Scene perception: Detecting and judging objects undergoing relational violations. Cogn. Psychol. 14 (2), 143–177.

Borkin, M., Bylinskii, Z., Kim, N., Bainbridge, C., Yeh, C., Borkin, D., Pfister, H., Oliva, A., 2016. Beyond memorability: Visualization recognition and recall. IEEE Trans. Vis. Comput. Graphics 22 (1), 519–528. http://dx.doi.org/10.1109/TVCG.2015.2467732.

Brumby, D., Zhuang, S., 2015. Visual grouping in menu interfaces. In: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, pp. 4203–4206. http://dx.doi.org/10.1145/2702123.2702177.

Buscher, G., Cutrell, E., Morris, M., 2009. What do you see when you're surfing? Using eye tracking to predict salient regions of web pages. In: CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, pp. 21–30. http://dx.doi.org/10.1145/1518701.1518705.

Buswell, G., 1935. How People Look at Pictures: A Study of the Psychology and Perception in Art. University of Chicago Press.

Byrne, M., 1993. Using icons to find documents: Simplicity is critical. In: CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, pp. 446–453. http://dx.doi.org/10.1145/169059.169369.

Byrne, M., Anderson, J., Douglass, S., Matessa, M., 1999. Eye tracking the visual search of click-down menus. In: CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, pp. 402–409. http://dx.doi.org/10.1145/302979.303118.

Carroll, B., 2009. Enjoying Grant Grove on a cool, cloudy spring day. URL https://flic.kr/p/HiH31o. Online image.

Chen, X., Zelinsky, G., 2006. Real-world visual search is dominated by top-down guidance. Vis. Res. 46 (24), 4118–4133. http://dx.doi.org/10.1016/j.visres.2006.08.008.

Chun, M.C., Wolfe, J., 1996. Just say no: How are visual searches terminated when there is no target present? Cogn. Psychol. 30 (1), 39–78. http://dx.doi.org/10.1006/cogp.1996.0002.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, second ed. Lawrence Erlbaum Associates, Hillsdale, NJ.

Cornia, M., Baraldi, L., Serra, G., Cucchiara, R., 2018. Predicting human eye fixations via an LSTM-based saliency attentive model. IEEE Trans. Image Process. 27 (10), 5142–5154. http://dx.doi.org/10.1109/TIP.2018.2851672.

Deka, B., Huang, Z., Franzen, C., Hibschman, J., Afergan, D., Li, Y., Nichols, J., Kumar, R., 2017. Rico: A mobile app dataset for building data-driven design applications. In: UIST '17: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. pp. 845–854.

Fisher, D., Coury, B., Tengs, T., Duffy, S., 1989. Minimizing the time to search visual displays: The role of highlighting. Hum. Factors 31 (2), 167–182. http://dx.doi.org/10.1177/001872088903100206, PMID 2744770.

Fleetwood, M., Byrne, M., 2006. Modeling the visual search of displays: A revised ACT-R model of icon search based on eye-tracking data. Hum.– Comput. Interact. 21 (2), 153–197. http://dx.doi.org/10.1207/s15327051hci2102_1.

Fosco, C., Casser, V., Bedi, A., O'Donovan, P., Hertzmann, A., Bylinskii, Z., 2020. Predicting visual importance across graphic design types. In: UIST '20: Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, New York, pp. 249–260. http://dx.doi.org/10.1145/3379337.3415825.

Gazepoint, 2021. GP3 eye-tracker. URL https://www.gazept.com.

Gillies, S., et al., 2007–2024. Shapely: Manipulation and analysis of geometric objects. URL https://github.com/Toblerity/Shapely.

Grahame, M., Laberge, J., Scialfa, C., 2004. Age differences in search of Web pages: The effects of link size, link number, and clutter. Hum. Factors 46 (3), 385–398. http://dx.doi.org/10.1518/hfes.46.3.385.50404, PMID 15573540.

Halverson, T., Hornof, A., 2004. Local density guides visual search: Sparse groups are first and faster. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 48 (16), 1860–1864. http://dx.doi.org/10.1177/154193120404801615.

Halverson, T., Hornof, A., 2011. A computational model of "active vision" for visual search in human–computer interaction. Hum.– Comput. Interact. 26 (4), 285–314. http://dx.doi.org/10.1080/07370024.2011.625237.

Henderson, J., 2005. Introduction to real-world scene perception. Vis. Cogn. 12 (6), 849–851.

Henderson, J., Ferreira, F., 2004. Scene perception for psycholinguists. In: Henderson, J., Ferreira, F. (Eds.), The Interface of Language, Vision, and Action. Psychology Press, pp. 1–58.

Hornof, A., 2004. Cognitive strategies for the visual search of hierarchical computer displays. Hum.– Comput. Interact. 19 (3), 183–223. http://dx.doi.org/10.1207/s15327051hci1903_1.

Hotchkiss, G., 2014. Google's Golden Triangle – nine years later.

Hotchkiss, G., Alston, S., Edwards, G., 2005. Eye tracking study: An in depth look at interactions with Google using eye tracking methodology.

Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20 (11), 1254–1259. http://dx.doi.org/10.1109/34.730558.

Jiang, Y., Leiva, L., Houssel, P., Tavakoli, H., Kylmälä, J., Oulasvirta, A., 2023. Ueyes: Understanding visual saliency across user interface types. In: CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, http://dx.doi.org/10.1145/3544548.3581096.

Johnson, J., 2014. Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines, fourth ed. Elsevier, Boston, MA.

Jokinen, J., Sarcar, S., Oulasvirta, A., Silpasuwanchai, C., Wang, Z., Ren, X., 2017. Modelling learning of new keyboard layouts. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 4203–4215.

Jokinen, J., Wang, Z., Sarcar, S., Oulasvirta, A., Ren, X., 2020. Adaptive feature guidance: Modelling visual search with graphical layouts. Int. J. Hum.– Comput. Stud. 136, http://dx.doi.org/10.1016/j.ijhcs.2019.102376.

Just, M., Carpenter, P., 1976. Eye fixations and cognitive processes. Cogn. Psychol. 8 (4), 441–480.

Kieras, D., Hornof, A., 2014. Towards accurate and practical predictive models of active-vision-based visual search. In: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, pp. 3875–3884.

Kümmerer, M., Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., Torralba, A., 2024. MIT/Tübingen Saliency Benchmark, https://saliency.tuebingen.ai/.

Kümmerer, M., Theis, L., Bethge, M., 2015. Deep gaze I: Boosting saliency prediction with feature maps trained on ImageNet. arXiv:1411.1045.

Kuznetsova, A., Brockhoff, P., Christensen, R., 2017. lmerTest package: Tests in linear mixed effects models. J. Stat. Softw. 82 (13), 1–26. http://dx.doi.org/10.18637/jss.v082.i13.

Leiva, L., Xue, Y., Bansal, A., Tavakoli, H., Köroðlu, T., Du, J., Dayama, N., Oulasvirta, A., 2020. Understanding visual saliency in mobile user interfaces. In: MobileHCI '20: 22nd International Conference on Human–Computer Interaction with Mobile Devices and Services. Association for Computing Machinery, New York, http://dx.doi.org/10.1145/3379503.3403557.

Levine, A., 2008. Kitchen. URL https://flic.kr/p/5yq5JS. Online image.

Ling, J., van Schaik, P., 2007. The influence of line spacing and text alignment on visual search of web pages. Displays 28 (2), 60–67. http://dx.doi.org/10.1016/j.displa.2007.04.003.

Liu, W., Cao, Y., Proctor, R., 2021. How do app icon color and border shape influence visual search efficiency and user experience? Evidence from an eye-tracking study. Int. J. Ind. Ergon. 84, http://dx.doi.org/10.1016/j.ergon.2021.103160.

Lüdecke, D., 2023. Sjplot: Data visualization for statistics in social science. URL https://CRAN.R-project.org/package=sjPlot. R package version 2.8.15.

Lüdecke, D., Ben-Shachar, M.S., Patil, I., Makowski, D., 2020. Extracting, computing and exploring the parameters of statistical models using R.. J. Open Source Softw. 5 (53), 2445. http://dx.doi.org/10.21105/joss.02445.

Miniukovich, A., De Angeli, A., 2015. Computation of interface aesthetics. In: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, pp. 1163–1172.
Miniukovich, A., Marchese, M., 2020. Relationship between visual complexity and aesthetics of webpages. In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–13.
Miniukovich, A., Sulpizio, S., De Angeli, A., 2018. Visual complexity of graphical user interfaces. In: AVI '18: Proceedings of the 2018 International Conference on Advanced Visual Interfaces.
Myers, C., Lewis, R., Howes, A., 2013. Bounded optimal state estimation and control in visual search: Explaining distractor ratio effects. Proc. Annu. Meet. Cogn. Sci. Soc. 35, 1049–1054.
Neider, M., Zelinsky, G., 2008. Exploring set size effects in scenes: Identifying the objects of search. Vis. Cogn. 16 (1), 1–10. http://dx.doi.org/10.1080/13506280701381691.
Nielsen, J., 2006. F-shaped pattern for reading web content (original study).
Nieminen, P., 2022. Application of standardized regression coefficient in meta-analysis. BioMedInformatics 2 (3), 434–458. http://dx.doi.org/10.3390/biomedinformatics2030028.
O'Donovan, P., Agarwala, A., Hertzmann, A., 2014. Learning layouts for single-page graphic designs. IEEE Trans. Vis. Comput. Graphics 20 (8), 1200–1213. http://dx.doi.org/10.1109/TVCG.2014.48.
Oulasvirta, A., De Pascale, S., Koch, J., Langerak, T., Jokinen, J., Todi, K., Laine, M., Kristhombuge, M., Zhu, Y., Miniukovich, A., Palmas, G., Weinkauf, T., 2018. Aalto interface metrics (AIM): A service and codebase for computational GUI evaluation. In: UIST '18 Adjunct: Adjunct Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, New York, pp. 16–19. http://dx.doi.org/10.1145/3266037.3266087.
Palmer, J., 1995. Attention in visual search: Distinguishing four causes of a set-size effect. Curr. Dir. Psychol. Sci. 4 (4), 118–123. http://dx.doi.org/10.1111/1467-8721.ep10772534.
Pernice, K., 2019. Text scanning patterns.
Pfeuffer, K., Li, Y., 2018. Analysis and modeling of grid performance on touchscreen mobile devices. In: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.
Resnick, M., Albert, W., 2014. The impact of advertising location and user task on the emergence of banner ad blindness: An eye-tracking study. Int. J. Hum.– Comput. Interact. 30 (3), 206–219. http://dx.doi.org/10.1080/10447318.2013.847762.
Resnick, M., Albert, W., 2016. The influences of design esthetic, site relevancy and task relevancy on attention to banner advertising. Interact. Comput. 28 (5), 680–694. http://dx.doi.org/10.1093/iwc/iwv042.
Rosenholtz, R., Li, Y., Nakano, L., 2007. Measuring visual clutter. J. Vis. 7 (2), http://dx.doi.org/10.1167/7.2.17.
Roth, S.P., Schmutz, P., Pauwels, S.L., Bargas-Avila, J.A., Opwis, K., 2010. Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages? Interact. Comput. 22 (2), 140–152. http://dx.doi.org/10.1016/j.intcom.2009.10.004, URL https://www.sciencedirect.com/science/article/pii/S0953543809000885.
Shen, J., Reingold, E., Pomplun, M., 2000. Distractor ratio influences patterns of eye movements during visual search. Perception 29 (2), 241–250.
Shen, C., Zhao, Q., 2014. Webpage saliency. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International, Cham, Switzerland, pp. 33–46.
Teo, L.-H., John, B., Blackmon, M., 2012. CogTool-explorer: A model of goal-directed user exploration that considers information layout. In: CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 2479–2488.
Tkachenko, M., Malyuk, M., Holmanyuk, A., Liubimov, N., 2020–2022. Label Studio: Data labeling software. URL https://github.com/heartexlabs/label-studio. Open source software available via https://github.com/heartexlabs/label-studio.
Todi, K., Jokinen, J., Luyten, K., Oulasvirta, A., 2019. Individualising graphical layouts with predictive visual search models. ACM Trans. Interact. Intell. Syst. 10 (1), http://dx.doi.org/10.1145/3241381.
Trapp, A., Wienrich, C., 2018. App icon similarity and its impact on visual search efficiency on mobile touch devices. Cogn. Res.: Princ. Implic. 3 (1).
Treisman, A., Gelade, G., 1980. A feature-integration theory of attention. Cogn. Psychol. 12 (1), 97–136.
Williams, L., 1967. The effects of target specification on objects fixated during visual search. Acta Psychol. 27, 355–360.
Wolfe, J., 1994. Visual search in continuous, naturalistic stimuli. Vis. Res. 34 (9), 1187–1195. http://dx.doi.org/10.1016/0042-6989(94)90300-X.
Wolfe, J., 2007. Guided search 4.0: Current progress with a model of visual search. In: Gray, W. (Ed.), Integrated Models of Cognitive Systems. pp. 99–119.
Wolfe, J., 2018. Visual search. In: Wixted, J. (Ed.), Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Volume 2: Sensation, Perception and Attention. John Wiley & Sons, Hoboken, NJ, pp. 569–623. http://dx.doi.org/10.1002/9781119170174.epcn213.
Wolfe, J., 2020. Visual search: How do we find what we are looking for? Annu. Rev. Vis. Sci. 6 (1), 539–562. http://dx.doi.org/10.1146/annurev-vision-091718-015048, PMID 32320631.
Wolfe, J., 2021. Guided search 6.0: An updated model of visual search. Psychon. Bull. Rev. 28 (4), 1060–1092.
Wolfe, J., Alvarez, G., Rosenholtz, R., Kuzmova, Y., Sherman, A., 2011a. Visual search for arbitrary objects in real scenes. Atten. Percept. Psychophys. 73, 1650–1671.
Wolfe, J., Horowitz, T., 2004. What attributes guide the deployment of visual attention and how do they do it? Nature Rev. Neurosci. 5 (6), 495–501.
Wolfe, J., Horowitz, T., Kenner, N., Hyle, M., Vasan, N., 2004. How fast can you change your mind? The speed of top-down guidance in visual search. Vis. Res. 44 (12), 1411–1426. http://dx.doi.org/10.1016/j.visres.2003.11.024.
Wolfe, J., Palmer, E., Horowitz, T., 2010. Reaction time distributions constrain models of visual search. Vis. Res. 50 (14), 1304–1311. http://dx.doi.org/10.1016/j.visres.2009.11.002.
Wolfe, J., Võ, M.L.-H., Evans, K., Greene, M., 2011b. Visual search in scenes involves selective and nonselective pathways. Trends Cogn. Sci. 15 (2), 77–84. http://dx.doi.org/10.1016/j.tics.2010.12.001.
Xie, M., Feng, S., Xing, Z., Chen, J., Chen, C., 2020. UIED: A hybrid tool for GUI element detection. In: ESEC/FSE 2020: Proceedings of the 28th ACM Joint Meeting – European Software Engineering Conference and Symposium on the Foundations of Software Engineering. Association for Computing Machinery, New York, pp. 1655–1659. http://dx.doi.org/10.1145/3368089.3417940.
Yarbus, A., 1967. Eye movements during perception of complex objects. In: Eye Movements and Vision. Springer, Boston, MA, pp. 171–211. http://dx.doi.org/10.1007/978-1-4899-5379-7_8.
Yuan, A., Li, Y., 2020. Modeling human visual search performance on realistic webpages using analytical and deep learning methods. In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, http://dx.doi.org/10.1145/3313831.3376870.
Zelinsky, G., 1996. Using eye saccades to assess the selectivity of search movements. Vis. Res. 36 (14), 2177–2187. http://dx.doi.org/10.1016/0042-6989(95)00300-2.