Size Does Matter: Effects of Icon Size Variation on Spatial Learning in Interfaces with Large Command Sets

Tahsin Aziz
University of Regina
Regina, Saskatchewan, Canada
taj541@uregina.ca

Md. Sami Uddin
University of Regina
Regina, Saskatchewan, Canada
sami.uddin@uregina.ca

Abstract

Effective command selection in graphical user interfaces (GUIs) depends on users' ability to remember command locations. While spatial memory-based interfaces can aid learning, the large number of commands in desktop applications complicates this process. Adding elements like coloured regions or background images can help, but they may also clutter interfaces and hinder learning. We investigate whether variation in icon size — an inherent GUI feature — can aid spatial learning. Therefore, we conducted two studies with varying icon sizes (small, large, and mixed) in interfaces with over 130 commands with meaningful and abstract icons, respectively. Results suggested that size variation, particularly enlarged icons, can improve performance and reduce errors. We show that variation in icon size can serve as an intrinsic spatial cue, thereby improving the design of spatial memory-based interfaces.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Command selection, landmark, menu design, spatial memory

ACM Reference Format:
Tahsin Aziz and Md. Sami Uddin. 2026. Size Does Matter: Effects of Icon Size Variation on Spatial Learning in Interfaces with Large Command Sets. In *Proceedings of Conference on Graphics Interface (GI '26).* ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Command selection is a fundamental operation in graphical user interfaces (GUIs), underlying nearly all interaction in productivity, visualization, and design software. Users invoke commands by locating and selecting items from menus, ribbons, or tool palettes, typically through visual search for novices or memory-based recall for experienced users [8, 14]. As modern applications often expose hundreds of commands, efficient interaction depends not only on pointing performance but on how quickly users can *learn and remember* command locations for repeated use [45, 54]. When spatial learning fails, users must resort to slow visual search, which can increase cognitive load and limit the development of expert performance [18, 23].

To facilitate learning and enable fast recall-based command selection, researchers explored spatially stable interfaces. Techniques such as CommandMaps [45], FastTap [23], and ListMaps [22] arrange commands in persistent spatial layouts to promote memory-driven interaction by forming object-location associations [14]. While effective for moderate command sets, these approaches often degrade as command sets scale [1, 54]. Repeated grids tend to be visually uniform, providing few distinctive reference points. Without salient anchors, which are essential for spatial learning, users may struggle to encode reliable spatial maps, slowing the transition from recognition-based search to recall-based selection [48, 53].

Spatial learning in physical environments relies on landmarks — distinct features like edges, corners, or notable objects that anchor cognitive maps and facilitate navigation [27, 38, 44]. Similar mechanisms apply in digital environments, where users rely on consistent layouts and reference points to organize spatial knowledge [57]. In GUIs, 'artificial landmarks' such as background images, colour blocks, or decorative elements have been introduced to support spatial learning [5, 6, 54]. Although effective in some cases, such extrinsic cues can lead to visual clutter and reduce usability as interface density increases [4]. This trade-off underscores the need for lightweight alternatives that can enhance spatial learning without introducing new visual elements.

In this paper, we explore leveraging an intrinsic property of GUIs—*command icon size*—as a spatial learning aid (i.e., landmark) for commands. Size is a fundamental visual variable, processed preattentively and capable of conveying hierarchy and distinctiveness with minimal visual overhead [10, 50, 60]. Gestalt principles further suggest that size differences can offer perceptual organization and grouping [29]. Unlike external embellishments, size variation integrates seamlessly into existing iconography and is already used in commercial GUIs as a design property to emphasize important commands. However, little is known about its role as a systematic cue for spatial learning and recall in dense interfaces.

We, therefore, conducted two controlled studies using large-scale, spatially stable desktop interfaces with 135–160 commands. Study 1 examined how variation in icon size (with large, small, and mixed sizes) influences learning and recall with semantically meaningful icons. Study 2 removed semantic cues by using abstract and uniform shapes to isolate spatial learning independent of semantic guidance. In two studies, we discovered that varying icon sizes could improve spatial learning. For semantic icons, different sizes enhanced performance only in later blocks (recall blocks; Study 1). Whereas for abstract icons, the benefits were consistent across all blocks (Study 2), and larger icons offered the greatest advantage.

GI '26, Kitchener-Waterloo, Ontario

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

This work makes three contributions. First, it highlights an underexplored aspect of an existing GUI feature, *icon size*, and investigates its potential as an intrinsic landmarking strategy for spatial learning of commands. Second, it provides empirical evidence—across semantic and abstract contexts— suggesting that variation in icon size can aid spatial learning and enhance recall-driven interactions without increasing workload. Third, it extends landmark-based models of spatial cognition to large-scale graphical interfaces, offering practical guidance for designing memory-supportive systems while maintaining visual simplicity.

2 Related Work

2.1 Spatial Memory and Command Selection

Efficient command selection depends on both pointing time and search time, where search decreases as users become familiar with command locations [14]. Novices primarily rely on visual search, whereas experienced users increasingly recall locations to enable faster selection [8, 25]. This shift is supported by spatial memory and object–location association learning [17, 18, 43, 49].

A range of techniques explicitly leverage spatial memory by stabilizing command locations. Pie menus [11] and marking menus [30] use consistent spatial structure to support expert recall-based selection. CommandMaps [45], ListMaps [22], and FastTap [23] present commands in spatially stable layouts that reduce hierarchical navigation and promote learning through repetition. Keyboard shortcuts can also improve efficiency, but they are constrained by learnability and limited command capacity [37, 40, 46].

However, maintaining spatial-memory benefits at large command scales remains challenging. Extensions of spatial overlays show that performance advantages can diminish as command capacity grows, motivating mechanisms that better support location learning in dense layouts [2, 4, 54].

2.2 Landmarks in Graphical Interfaces

Landmarks provide reference points that support orientation and spatial memory in both physical and digital environments [35, 53]. In GUIs, edges and corners often function as natural landmarks that users exploit to remember nearby targets [51, 52]. Many small-screen techniques capitalize on these cues by placing controls near boundaries [23, 31, 47], but on large displays, substantial interface regions lie far from boundary anchors, limiting their usefulness.

To compensate, prior work explored additional landmarks, such as embodied reference frames (e.g., hands) [56], and artificial visual cues, including colour and imagery [5, 33]. Artificial landmarks have been applied to revisitation and navigation support [5, 26, 55], command selection in dense interfaces [3, 20, 54], and other interaction contexts [21, 34, 36, 41, 42]. Yet, in spatially stable 2D command layouts, the design space of landmarks that remain effective at scale without introducing clutter is still limited. Prior augmentations of CommandMaps and related overlays indicate that feature-rich image cues can underperform compared to simpler cues such as colour blocks, suggesting a trade-off between memorability and visual clutter [4, 54]. This motivates investigating intrinsic GUI features that can act as landmarks without adding new visual layers.

2.3 Icon Size Variation as an Intrinsic Landmark

In visualization and interface design, *size* is a fundamental visual variable that can guide attention and encode structure [10, 58]. Along with other visual variables, size differences can be discriminated rapidly and can support efficient selection behaviour [7, 24, 50, 61]. HCI work has used size variation to improve performance in recognition-heavy tasks and to influence selection efficiency [14, 32], while perceptual organization principles (e.g., Gestalt) help explain how such variation can structure visual grouping [58].

Critically, recognition and recall have different cognitive requirements. As users gain expertise, they transition from search-dominant behaviour to memory-driven retrieval, consistent with skill-acquisition accounts [19]. While prior work has examined how visual variables support recognition, the role of size variation in supporting recall and spatial learning in dense, spatially stable command grids is less understood. We examine whether icon size variation can act as an intrinsic landmark that supports spatial learning and recall without the clutter costs of extrinsic landmarks.

3 Study 1: Icon Size Effects with Semantic Icons

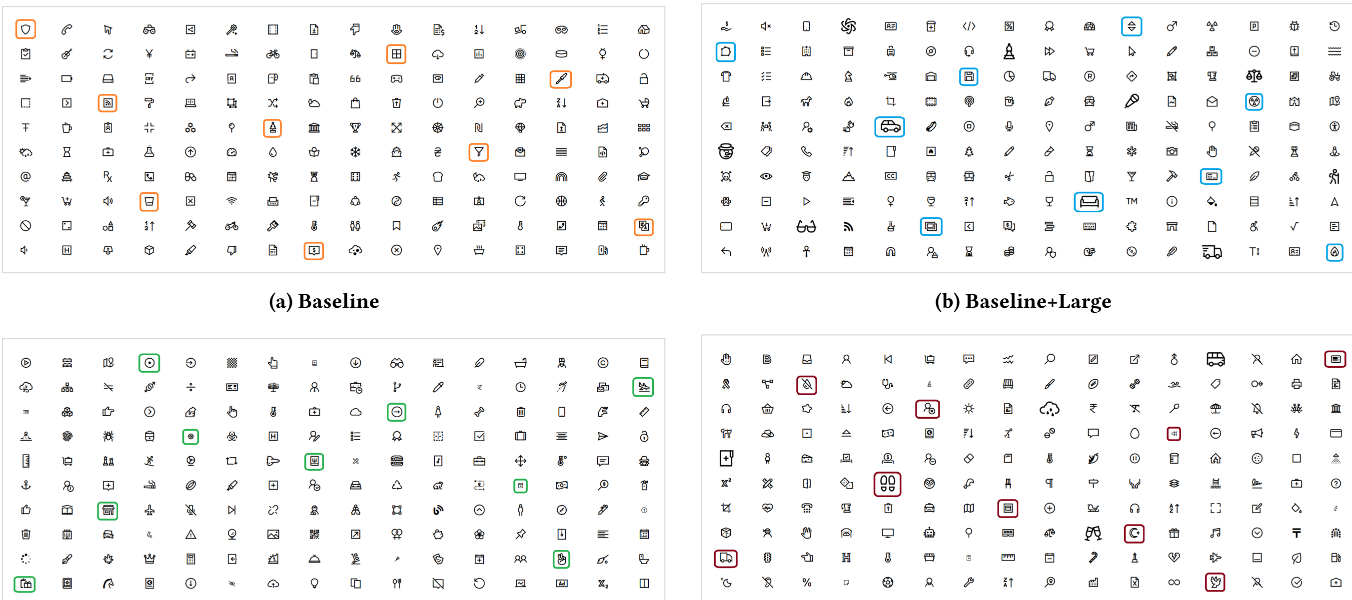
Study 1 investigates whether icon size variation can function as an *intrinsic visual landmark* that may support spatial learning and recall in large, spatially stable command interfaces. The study focuses on the transition from search-driven novice behaviour to memory-driven expert performance in dense GUIs [9, 22].

3.1 Study 1 Interfaces

We designed four CommandMaps-style interfaces [45] displaying 160 grayscale, semantically meaningful icons from the *Line Awesome* set [28] arranged in a fixed 10×16 grid. The grid occupied nearly the entire display, except for a 140 px region at the bottom reserved for target stimuli. Icons were randomly assigned to interfaces to avoid potential grouping effects; no icon was reused in other interfaces. The baseline icon size was 45 px. The *Baseline* interface used a uniform size. In *Baseline+Large*, 10 icons were enlarged to 80 px; in *Baseline+Small*, 10 icons were reduced to 25 px; and *Baseline+Mixed* combined 5 large (80 px) and 5 small (25 px) icons. Size-modified icons were distributed across the grid (unique to each interface) to avoid clustering, introducing perceptual distinctiveness while preserving grid regularity and icon recognizability in a dense layout. To prevent confounding pointing performance, clickable regions were identical across all icons. Example target layouts are shown in Figure 1.

3.2 Study 1 Method

3.2.1 Task and Procedure. Participants performed repeated target-selection tasks. On each trial, a target label appeared at the bottom of the display, and participants selected the corresponding icon using a mouse. Correct selections were confirmed by highlighting the selected cell in green for 500 ms, then removing all icons from the display; incorrect ones in red, and participants could proceed only after selecting the correct target. To reset the cursor position and keep the starting position of each trial uniform, we added a "Next" button at the bottom of the screen; clicking it revealed the next target and icon grid. They were instructed to complete the tasks quickly and accurately, with allowed pauses between trials



(a) Baseline

(b) Baseline+Large

(c) Baseline+Small

(d) Baseline+Mixed

Figure 1: Interface variants used in the Study 1. Targets used as stimuli are shown in colour boxes.

to reduce fatigue. Cursor position was not explicitly controlled at the start of each trial, which may have introduced variability in movement distance. However, because target positions were counterbalanced across interfaces and participants had to click the "Next" button to proceed, any such effects are unlikely to explain the observed interface-level differences.

Each interface comprised 12 blocks of 9 trials. A set of 9 unique targets per interface was used, repeated across blocks in a randomized order. The targets selected from edges or sides (2), corners (2), and central regions of the grid (5). Central targets included items located on landmarks (2) as well as near (1 cell away: 2) and far (2 or more cells away: 1) from them. Target positions varied across the interfaces to avoid potential learning effects. Before the study, participants completed practice trials using a separate icon set to familiarize themselves with the task and feedback.

Participants were informed that the study examined interfaces for rapid command selection. After providing informed consent, they completed a demographic questionnaire and a Corsi Block Test [16], then received task instructions and were asked to perform the task as quickly and accurately as possible. The study used a within-participant design, with interface order counterbalanced using a balanced Latin square. After completing each interface, participants filled out a NASA–TLX questionnaire and concluded the session with a brief interview about interface preferences.

Following prior work on learning dynamics in spatially stable interfaces [8, 14], Blocks 1–3 were treated as *learning* blocks and Blocks 4–12 as *recall* blocks. Here, *recall* refers to later-session performance characterized by increased reliance on learned object–location associations while the interface remains visible.

3.2.2 Participants and Apparatus. Sixteen participants (8 women, 8 men), aged 24–47 ($M = 31.5$), with normal or corrected-to-normal vision, were recruited from a university population. Sessions (conducted in a quiet indoor setting) lasted approximately 60 minutes; participants received CA$10 compensation. The experiment ran on a Windows 11 desktop with a 27-inch monitor (1920×1080). At an approximate viewing distance of 60 cm, 25 px, 45 px, and 80 px icons subtended roughly 0.7°, 1.2°, and 2.1° of visual angle, respectively. Data processing and statistical analyses were conducted in *R*. The University of Regina Research Ethics Board approved the study protocol.

3.3 Study 1 Results

Shapiro–Wilk tests and Q–Q plot inspections indicated approximately normal residual distributions (all $W > 0.9$, $p > 0.05$); so we used repeated-measures ANOVA for inferential analyses over all blocks and separately for learning and recall phases. We report generalized eta-squared (η^2) [15] with thresholds as .01 (small), .06 (medium), and > .14 (large). Post-hoc tests used TukeyHSD. Most participants (14/16) had a Corsi span of 6, with only two deviations (8 and 4); therefore, we analyzed participants as a single group. Performance improvements across blocks were consistent with the typical power-law learning ($R^2 > .92$) pattern observed in repeated selection tasks.

3.3.1 Completion Time Analysis. Completion time was measured from the onset of the target cue to the correct selection. Mean completion times are summarized in Table 1 and in Figure 2.

Overall, completion times decreased sharply over time across all interfaces, yielding a significant main effect of Block ($F_{11,165} = 199.43$, $p < .001$, $\eta^2 = .77$) in the ANOVA. There was no main

Table 1: Study 1 completion time (ms) by block phases with standard deviations (s.d.).

Interface	Blocks		
	All	Learning	Recall
Baseline	5216.06 (11512.09)	11511.56 (19081.71)	3117.57 (6158.49)
Baseline+Large	4628.28 (10904.43)	11814.25 (19451.10)	2232.96 (3111.99)
Baseline+Small	5062.10 (15222.87)	13865.65 (28308.93)	2127.58 (2801.63)
Baseline+Mixed	4801.50 (12048.66)	11449.07 (21627.51)	2585.65 (4273.96)

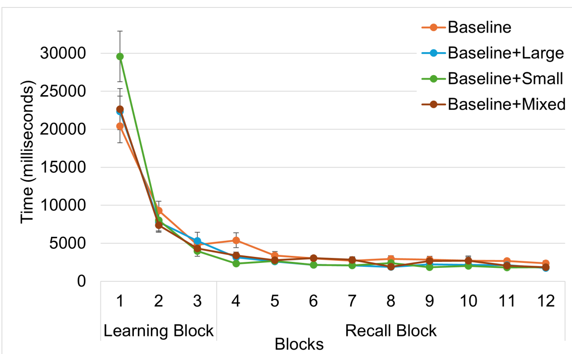


Figure 2: Mean completion time (s.e.) by blocks for Study 1.

effect of Interface across all blocks ($F_{3,45} = 0.55$, $p = .65$), indicating no reliable overall difference when averaging across learning and recall. However, a significant Interface×Block interaction ($F_{33,495} = 2.74$, $p < .001$, $\eta^2 = .10$) indicates that interfaces differed in how performance evolved over time.

When analyzing phases separately, learning blocks (Blocks 1–3) showed no reliable main effect of Interface ($F_{3,45} = 1.40$, $p = .25$), despite a significant Interface×Block interaction ($F_{6,90} = 3.04$, $p = .009$). This suggests that early performance was driven by general familiarity rather than by layout differences.

In contrast, during recall blocks (Blocks 4–12), a significant main effect of Interface emerged ($F_{3,45} = 3.44$, $p = .024$, $\eta^2 = .06$). A significant effect of block was also found ($F_{8,120}$=6.32, p<0.001, η^2=0.08). However, no significant Interface× Block interaction was observed ($F_{24,360}$=1.41, p=0.0986) as performance stabilized.

Post-hoc tests showed that the Baseline condition was significantly slower than the other interfaces (all p<0.05). Among the varied icon-size conditions, Baseline+Small yielded faster completion time than Baseline+Mixed (p<0.05). These results indicate that the effects of icon size variation were neutral in the early blocks but contributed to faster performance in the later blocks.

3.3.2 Error Analysis. For each trial, incorrect selections were treated as errors. Mean errors are reported in Table 2 and in Figure 3.

Table 2: Study 1 mean errors (s.d.) by block phases.

Interface	Blocks		
	All	Learning	Recall
Baseline	0.02 (0.14)	0.02 (0.17)	0.01 (0.13)
Baseline+Large	0.02 (0.19)	0.07 (0.35)	0.01 (0.08)
Baseline+Small	0.02 (0.20)	0.05 (0.35)	0.01 (0.12)
Baseline+Mixed	0.02 (0.27)	0.06 (0.52)	0.01 (0.09)

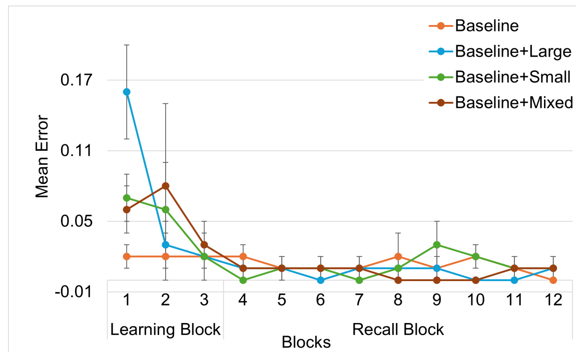


Figure 3: Mean errors (s.e.) across blocks for Study 1.

Overall errors were low across all interfaces and declined over blocks, producing a significant main effect of *Block* ($F_{11,165} = 4.71$, $p < .001$, $\eta^2 = .10$). However, we found no effect of *Interface* ($F_{3,45} = 1.06$, $p = .37$), though a significant *Interface*×*Block* interaction was observed ($F_{33,495} = 2.00$, $p = .001$, $\eta^2 = .075$).

However, for the learning phase, RM-ANOVA revealed significant effects of interface ($F_{3,45}$=2.87, p<0.05, η^2=0.03) and Interface × Block interaction ($F_{6,90}$= 2.21, p<0.05, η^2=0.07), but no effects of block ($F_{2,30}$=3.16). Post-hoc comparisons, however, showed no significant differences among any interface pairs (all p > 0.21).

In contrast, during the recall phase (Blocks 4–12), no significant effects were observed for *Interface* ($F_{3,45} = 1.20$, $p = .322$), *Block* ($F_{8,120} = 0.42$, $p = .905$), or their interaction ($F_{24,360} = 0.88$, $p = .63$).

Taken together, these results demonstrate that the faster completion times observed for size-variant interfaces were *not* accompanied by increased error rates. Performance improvements, therefore, reflect more efficient selection and stronger spatial learning.

3.3.3 Interface Position Analysis. Since interface corners and sides, or bezels, are already found to enhance spatial learning of nearby items [23, 53], we adjusted the icon sizes mainly in the central area of our GUIs and categorized trial questions into two groups: *Middle* and *Side+Corner*. Out of 160 icons, 88 were classified as *Side+Corner*(icons in corners, along the outer sides, and immediately adjacent) and 72 as *Middle*. For the trials, we selected four from the *Side+Corner* group (two from sides/corners and two adjacent) and five from the *Middle* group, maintaining a balanced ratio across both categories. Completion times by target position are summarized in Table 3 and Figure 4.

Table 3: Study 1 completion time by position (ms) with s.d.

Position	Blocks		
	All	Learning	Recall
Side+Corner	4131.87 (10273.08)	10639.96 (18749.58)	1962.51 (2194.79)
Middle	5563.08 (14050.70)	13376.27 (24926.96)	2958.69 (5397.94)

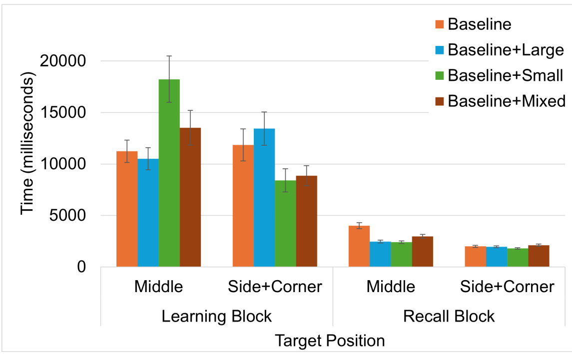


Figure 4: Study 1 completion time by target position (Side+Corner vs. Middle) across learning and recall phases.

Across all blocks, ANOVA revealed a significant effect of *Position*, indicating that targets located at the sides and corners were selected faster than those in the middle of the interface ($F_{1,15} = 35.35$, $p < .001$, $\eta^2 = .027$). The main effect of *Interface* was not significant ($F_{3,45} = 0.43$, $p = .732$). However, a significant *Interface*×*Position* interaction was observed ($F_{3,45} = 5.70$, $p = .002$, $\eta^2 = .019$), indicating that interface-related performance differences depended on target position rather than being uniform across the grid.

For learning blocks, a significant main effect of *Position* was observed ($F_{1,15} = 23.19$, $p < .001$, $\eta^2 = .031$), confirming faster selection for Side+Corner targets during early visual search. The effect of *Interface* was not significant ($F_{3,45} = 0.95$, $p = .425$). A significant *Interface*×*Position* interaction was present ($F_{3,45} = 11.74$, $p < .001$, $\eta^2 = .095$), indicating that interface differences during learning were position-dependent.

During recall blocks, the effect of *Position* remained significant ($F_{1,15} = 14.99$, $p = .002$, $\eta^2 = .068$), indicating a persistent advantage for Side+Corner targets. The main effect of *Interface* reached significance ($F_{3,45} = 3.29$, $p = .029$, $\eta^2 = .036$), and the *Interface*×*Position* interaction was also significant ($F_{3,45} = 3.89$, $p = .015$, $\eta^2 = .026$). Post-hoc tests showed that interface-related differences were confined to Middle targets: the Baseline was significantly slower than the size-variant interfaces (all $p < 0.001$), whereas Side+Corner targets showed no reliable differences between interfaces.

3.3.4 Landmark Proximity Analysis. To examine whether performance differences were influenced by targets located on size-modified icons, we analyzed completion times based on targets' proximity to landmarks and task phase (learning and recall blocks).

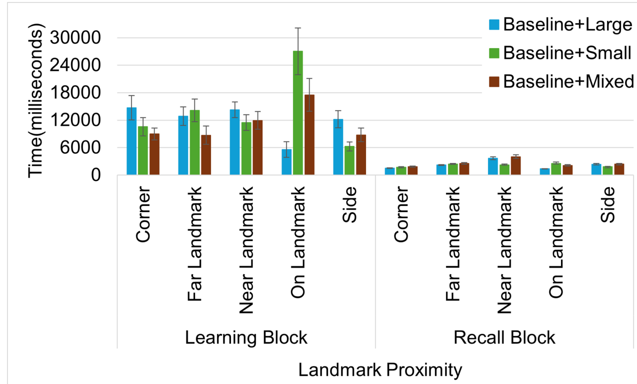


Figure 5: Study 1 completion time (ms) by landmark proximity across learning and recall blocks.

As shown in Figure 5, completion times varied significantly with landmark proximity. A repeated-measures ANOVA revealed a significant main effect of proximity in both learning ($F_{4,60} = 9.94, p < .001, \eta^2 = .037$) and recall phases ($F_{4,60} = 18.38, p < .001, \eta^2 = .059$). Post-hoc comparisons in the recall phase showed that targets near landmarks were significantly slower than targets at corner, far, and on-landmark positions (all $p < .001$).

During learning, differences across proximity conditions were less consistent, with no clear advantage for targets located on landmarks. In contrast, during recall, targets on landmarks showed improved performance comparable to corner positions, while near-landmark targets remained slower. This pattern suggests that performance differences were not solely driven by selecting size-modified icons, but instead reflect the development of spatial memory supported by size-based landmarks.

3.3.5 Subjective Workload and Preferences. Aligned Rank Transform [59] was used to perform RM-ANOVA on the NASA-TLX responses. No significant differences in workload were observed across interfaces, suggesting that introducing size variation did not measurably increase perceived effort, whereas objective performance differences emerged in later blocks (Table 4). Participant preference counts revealed that participants most frequently preferred *Baseline+Large* for speed, memorability, and overall experience, consistent with recall-phase performance results (Table 5).

4 Study 2: Icon Size Effects with Abstract Icons

In Study 1, participants engaged with meaningful icons, using their prior knowledge to aid spatial learning. Study 2, however, focused on abstract shapes without inherent meaning, requiring participants to rely solely on size and position cues. It helped us assess how icon-size variation affects learning independently of semantic memory.

4.1 Study 2 Interfaces

As in Study 1, three grid interfaces were designed, each containing 135 abstract shapes arranged in a fixed 9×15 grid (Figure 6). To maintain a similar session duration while retaining a large command set, the total number of items was reduced from 160 to 135. Each shape was assigned a unique five-character label as a temporary

Table 4: Study 1 mean NASA–TLX scores (0 = low, 7 = high). Standard deviations in parentheses.

Subscale	Interface				F	p
	Baseline	Baseline +Large	Baseline +Small	Baseline +Mixed		
Mental Demand	3.44 (2.00)	3.56 (1.63)	4.00 (1.75)	3.38 (1.31)	1.70	.181
Physical Demand	2.50 (1.55)	2.44 (1.55)	2.81 (1.72)	2.19 (1.38)	1.06	.375
Temporal Demand	2.50 (1.51)	2.69 (1.62)	3.12 (1.63)	2.75 (1.48)	2.35	.085
Performance (rev.)	4.88 (2.13)	5.31 (1.70)	5.38 (1.26)	4.75 (2.08)	0.53	.664
Effort	3.38 (1.75)	3.44 (1.59)	3.88 (1.63)	3.31 (1.62)	0.79	.505
Frustration	2.62 (1.89)	2.69 (1.96)	2.94 (1.48)	2.62 (1.45)	0.57	.636

Table 5: Study 1 participant preferences across interfaces.

Measure	Baseline	Baseline +Large	Baseline +Small	Baseline +Mixed
Speed	1	9	3	3
Accuracy	2	7	3	4
Memorization	0	9	2	5
Expertise	0	12	3	1
Comfort	2	6	5	3
Overall	1	8	2	5

visual cue, with no inherent semantic meaning. A 140 px bottom margin was reserved for the target cue. Each interface used a distinct abstract shape (oval for Baseline, square for Baseline+Large, and circle for Baseline+Small) to prevent learning transfer between conditions. Landmark and target positions were also varied across interfaces.

The baseline item size was 65 px. In the *Baseline* interface, all shapes were rendered uniformly. In *Baseline+Large*, seven shapes were enlarged to 90 px, and in *Baseline+Small*, seven shapes were reduced to 40 px. Size-modified shapes were distributed across the grid to avoid clustering. As in Study 1, clickable regions were identical across items to avoid confounding pointing difficulty. In addition, Study 2 introduces *blind blocks*, in which visual labels are disabled, requiring participants to rely solely on spatial memory.

We removed *Baseline+Mixed* from Study 2. Pilot feedback and prior visualization principles suggested that multiple size variations within a single layout increased visual complexity and could introduce unintended perceptual grouping. Study 2, therefore, focuses on isolating the effects of a single size contrast.

4.2 Study 2 Method

4.2.1 Task and Procedure. Participants performed repeated target-selection tasks similar to Study 1. In *regular blocks*, item labels were hidden by default and could be revealed only while holding the `Ctrl` key after a 700 ms delay; labels disappeared upon releasing `Ctrl`. Participants could also select items without revealing labels. This design encouraged a gradual transition from visually guided search toward memory-driven selection.

Each interface consisted of 12 blocks of 9 trials, with the same target set used in randomized order. Targets were distributed across six spatial categories: corner, side, on-landmark, near landmark (one cell away), far landmark (two cells away), and far (more than two cells away), comprising a total of nine target items per interface. Two blocks (Blocks 6 and 12) were *blind blocks* (Figure 7), in which label reveal was disabled. Participants made a single selection, recalling the location from memory, and the trial advanced immediately regardless of accuracy. This prevented the use of corrective search strategies and provided a strict measure of spatial recall. The blind blocks were strategically placed to assess recall at different learning stages. Block 6 occurred after several regular blocks, serving as an early test of memory-based selection, while Block 12, at the end of the condition, acted as a retention check. We limited the blind blocks to 2 to avoid frequent disruptions and prevent the task from devolving into guessing without feedback. The study used a within-participant design with interface order counter-balanced across participants.

4.2.2 Participants and Apparatus. Twenty-five new participants were recruited from the local university community through campus announcements. Two participants were excluded due to incomplete experimental sessions. Data from 23 participants (15 men, 8 women), aged 19–49 (M = 28.0), with normal or corrected-to-normal vision, were included after exclusions; participants received CA$10 or course credit. Participants were familiar with desktop GUIs and completed the study in approximately 60 minutes. The experimental setup was identical to that of Study 1.

4.3 Study 2 Results

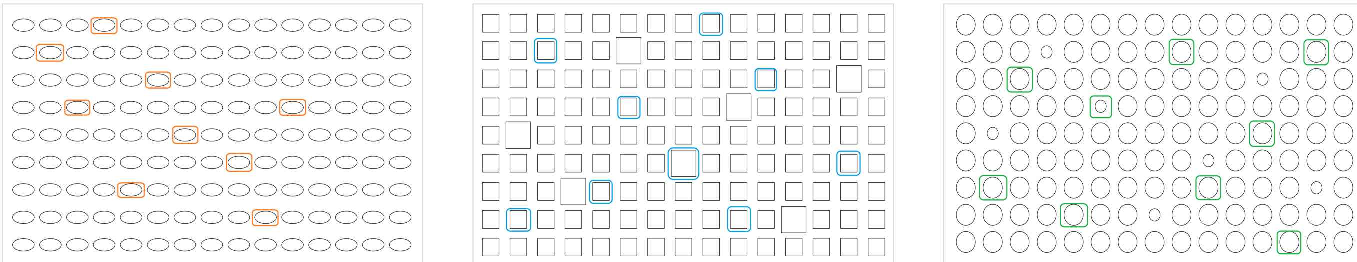
Shapiro–Wilk tests and Q–Q plots indicated approximately normal residuals ($W > 0.9$, $p > 0.05$); we therefore analyzed data using RM-ANOVAs similar to Study 1. Participants completed a Corsi Block Test; spans were typically 5–7 (20/23), with a few outliers, so we analyzed participants as a single group. Study 2 also included *blind blocks*, in which participants made selections from their memory; results were therefore analyzed separately for regular and blind blocks. To quantify reliance on the visual aid during regular blocks, we also analyzed the proportion of `Ctrl`-key use.

4.3.1 Completion Time Analysis. Mean completion times are summarized in Table 6 and in Figure 8.

Table 6: Study 2 completion time (ms) with s.d.

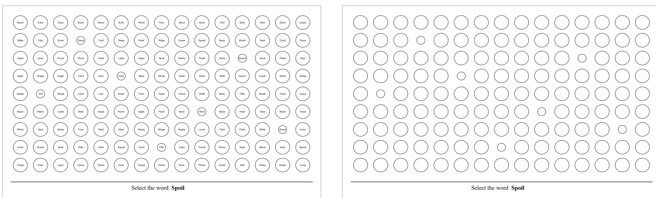
Interface	Regular Blocks	Blind Blocks
Baseline	12456.34 (20494.64)	3853.87 (3717.09)
Baseline+Large	8910.74 (17175.39)	2881.28 (3126.18)
Baseline+Small	9716.47 (19584.44)	2892.43 (3081.15)

For the regular blocks, completion times decreased over blocks across all interfaces, yielding a significant main effect of *Block*



(a) Baseline **(b) Baseline+Large** **(c) Baseline+Small**

Figure 6: Interface variants used in Study 2 with abstract shapes. Targets used as stimuli are shown in colour boxes.



(a) Regular block: labels revealed while holding `Ctrl`. **(b) Blind block: no labels; selection based on memory only.**

Figure 7: Selection modes in Study 2. In regular blocks, labels appear only after a short delay when holding `Ctrl`, aiding visual search. In blind blocks, label reveal is disabled, allowing only memory-based recall.

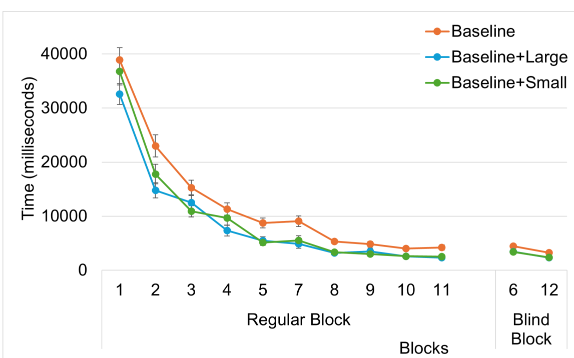


Figure 8: Mean completion time (s.e.) by blocks for Study 2.

($F_{9,198} = 108.11$, $p < .001$, $\eta^2 = .69$), following the expected power-law function ($R^2 > .98$) [39]. A significant main effect of *Interface* was also observed ($F_{2,44} = 9.95$, $p < .001$, $\eta^2 = .05$), indicating reliable completion-time differences between interfaces. However, the *Interface×Block* interaction was not significant ($F_{18,396} = 1.15$, $p = .298$, $\eta^2 = .02$), suggesting that interfaces differed primarily in overall performance level rather than in the rate of learning across blocks. Consequently, we avoided analyzing the performance by separating the blocks into learning and recall, as done in Study 1.

Post-hoc tests indicated Baseline was significantly slower than Baseline+Large and Baseline+Small (both $p < .001$), while the two interfaces with variable-sized icons performed identically (p=0.37).

During blind blocks, ANOVA again showed a significant main effect of *Interface* ($F_{2,44} = 10.40$, $p < .001$, $\eta^2 = .16$) and a significant main effect of *Block* ($F_{1,22} = 18.56$, $p < .001$, $\eta^2 = .23$), with no significant *Interface×Block* interaction ($F_{2,44} = 0.15$, $p = .858$). Tukey tests confirmed that both size-variant interfaces were significantly faster than the Baseline, while no reliable differences were observed between the two size-variant interfaces.

4.3.2 Error Analysis. Mean error rates are summarized in Table 7 and in Figure 9. Overall error rates were low across all interfaces, with modest increases during later regular blocks and blind blocks, reflecting the increased difficulty of recall-based selection.

Table 7: Study 2 mean errors with standard deviations.

Interface	Regular Blocks			Blind Blocks
	All	Learning	Recall	
Baseline	0.35 (2.81)	0.08 (0.57)	0.47 (3.33)	0.34 (0.47)
Baseline+Large	0.20 (2.40)	0.38 (3.63)	0.12 (1.60)	0.14 (0.34)
Baseline+Small	0.13 (0.62)	0.11 (0.51)	0.13 (0.67)	0.17 (0.38)

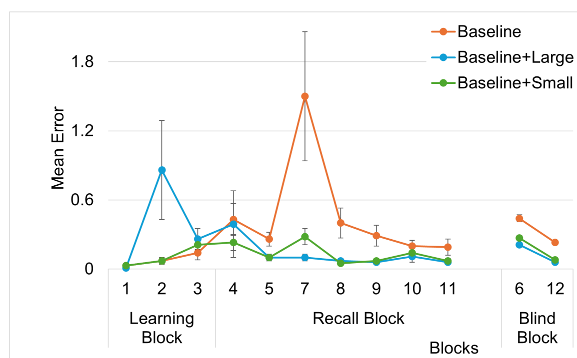


Figure 9: Mean errors (s.e.) across blocks for Study 2.

For the regular blocks, ANOVA revealed a significant effect of *Block* ($F_{9,198} = 2.11$, $p = .030$, $\eta^2 = .028$), indicating that error rates varied across blocks as participants progressed through the task. It showed no effect of *Interface* ($F_{2,44} = 1.58$, $p = .217$), indicating no reliable overall differences in accuracy when averaging across blocks. However, the *Interface×Block* interaction was significant ($F_{18,396} = 2.11$, $p = .005$, $\eta^2 = .052$).

For analysis, regular blocks were divided into *learning* (Blocks 1–3) and *recall* (Blocks 4–11, excluding blind blocks). When learning blocks were analyzed separately, no effects were significant (Interface: $F_{2,44} = 1.20$, $p = .312$; Block: $F_{2,44} = 1.51$, $p = .232$; Interface×Block: $F_{4,88} = 1.34$, $p = .261$), indicating comparable early accuracy across interfaces.

In contrast, during recall blocks, significant effects emerged for all factors. ANOVA showed a significant main effect of *Interface* ($F_{2,44} = 3.56$, $p = .037$, $\eta^2 = .031$), a significant main effect of *Block* ($F_{6,132} = 2.87$, $p = .012$, $\eta^2 = .036$), and a significant *Interface×Block* interaction ($F_{12,264} = 2.04$, $p = .021$, $\eta^2 = .046$). Post-hoc comparisons indicated that the Baseline produced significantly more errors than both size-variant interfaces (both p=0.0001), while the two size-variant interfaces did not differ reliably from each other. These results indicate that accuracy advantages for size-based landmarks emerged primarily during recall, when participants increasingly relied on learned object–location associations.

During blind blocks, error rates remained low overall but showed clear differences across interfaces. ANOVA revealed significant main effects of *Interface* ($F_{2,44} = 11.22$, $p < .001$, $\eta^2 = .233$) and *Block* ($F_{1,22} = 45.27$, $p < .001$, $\eta^2 = .250$), with no significant *Interface×Block* interaction ($F_{2,44} = 0.43$, $p = .654$, $\eta^2 = .005$). Tukey comparisons showed that the Baseline produced significantly more errors than both size-variant interfaces (both p<0.001), while no reliable differences were observed between the size-variant interfaces. The persistence of these accuracy differences without visual labels is consistent with more stable object–location memory in the size-variant interfaces rather than a speed–accuracy trade-off.

4.3.3 Interface Position Analysis. The interface contained 48 Side+Corner items and 87 Middle items; to emphasize the anchor-poor central region while maintaining representation, targets included 2 Side+Corner and 7 Middle items per interface.

Completion times by target position are summarized in Table 8 and Figure 10. Across phases, targets located at the sides and corners were selected faster than targets in the middle of the interface, indicating a robust advantage for positions near natural anchors.

Table 8: Study 2 completion time by position (ms) with s.d.

Position	Regular Blocks			Blind Blocks
	All	Learning	Recall	
Side+Corner	7836.99 (17126.08)	18754.18 (26088.78)	3109.25 (7215.66)	2599.09 (2182.04)
Middle	11075.67 (19681.92)	23556.74 (28388.75)	5742.42 (10662.08)	3382.70 (3595.12)

ANOVA across all regular blocks revealed a significant main effect of *Position* ($F_{1,22} = 24.75$, $p < .001$, $\eta^2 = .2317$), confirming that Side+Corner targets were completed faster than Middle targets. A significant main effect of *Interface* was also observed ($F_{2,44} = 5.80$, $p = .0058$, $\eta^2 = .1130$), along with a significant *Interface×Position* interaction ($F_{2,44} = 10.58$, $p < .001$, $\eta^2 = .1066$), indicating that

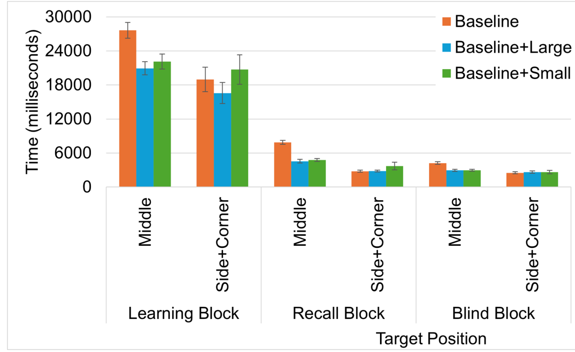


Figure 10: Study 2 completion time by target position (Side+Corner vs. Middle) across regular learning, regular recall, and blind blocks.

interface-related differences depended on target position rather than being uniform across the grid.

When analyzing phases separately, learning blocks showed significant main effects of *Interface* ($F_{2,44} = 3.94$, $p = .0267$, $\eta^2 = .0647$) and *Position* ($F_{1,22} = 11.28$, $p = .0028$, $\eta^2 = .1027$), while the *Interface×Position* interaction did not reach significance ($F_{2,44} = 2.49$, $p = .0948$). Follow-up tests indicated that these interface differences were driven primarily by *Middle* targets: the Baseline was slower than both size-variant interfaces for Middle targets (both $p < .007$), whereas no reliable pairwise differences were observed for Side+Corner targets.

In recall blocks, the main effect of *Position* remained significant ($F_{1,22} = 26.56$, $p < .001$, $\eta^2 = .1953$), and the *Interface×Position* interaction was significant ($F_{2,44} = 10.35$, $p < .001$, $\eta^2 = .0976$), while the main effect of *Interface* did not reach significance ($F_{2,44} = 2.55$, $p = .0895$, $\eta^2 = .0619$). Tukey follow-ups again showed that interface differences were confined to Middle targets: the Baseline was slower than both size-variant interfaces for Middle targets (all $p < .0001$), while Side+Corner targets did not show reliable interface differences. This pattern indicates that size-based landmarks selectively reduced the middle-region disadvantage when performance depended more strongly on recall.

A similar pattern was observed during blind blocks. ANOVA revealed significant main effects of *Interface* ($F_{2,44} = 4.62$, $p = .015$, $\eta^2 = .0752$) and *Position* ($F_{1,22} = 23.84$, $p < .001$, $\eta^2 = .1471$), as well as a significant *Interface×Position* interaction ($F_{2,44} = 5.88$, $p = .005$, $\eta^2 = .1081$). Follow-up comparisons again indicated that differences were concentrated in the Middle region, with the Baseline slower than both size-variant interfaces for Middle targets (both $p < .0001$) and no reliable differences among interfaces for Side+Corner targets.

Together, these results replicate the robust advantage of edges and corners as natural spatial reference regions, while showing that size-based landmarks selectively mitigate the middle-region disadvantage during recall and blind blocks, when participants rely more strongly on learned object–location associations rather than immediate visual guidance.

4.3.4 Landmark Proximity Analysis. We further examined whether performance varied with proximity to size-modified landmarks.

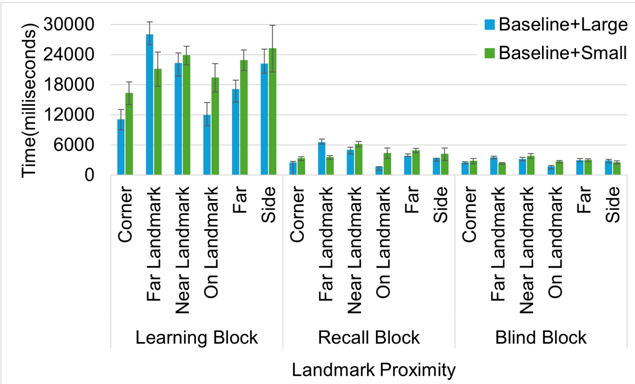


Figure 11: Study 2 completion time (ms) by landmark proximity across learning, recall, and blind blocks.

As shown in Figure 11, completion times varied across proximity conditions. A repeated-measures ANOVA revealed a significant main effect of landmark proximity in learning ($F_{5,110} = 6.33, p < .001, \eta^2 = .131$), recall ($F_{5,110} = 4.08, p = .002, \eta^2 = .066$), and blind blocks ($F_{5,110} = 2.87, p = .018, \eta^2 = .057$).

Post-hoc comparisons indicated that targets near landmarks were slower than targets at corner and on-landmark positions, particularly in recall and blind blocks (all $p < .05$).

Across blocks, targets near landmarks consistently exhibited higher completion times, while targets on landmarks and at corner positions tended to show relatively lower completion times. However, this pattern was more pronounced in recall and blind blocks than in the learning phase.

Importantly, similar performance patterns were observed during blind blocks, when visual labels were unavailable. This suggests reduced reliance on immediate visual cues and is consistent with the use of spatial memory, indicating that size-based landmarks may support memory-based representations of the interface.

4.3.5 Visual Aid Use. Use of the visual aid (Ctrl key) was analyzed as an indicator of reliance on visual confirmation versus memory-driven selection. Ctrl-key use was analyzed using a GLMM (binomial link) with participant as a random intercept to account for repeated measures. The analysis included one observation per trial, yielding $23 \times 3 \times 12 \times 9 = 7{,}452$ binary observations (participant $\times$ interface $\times$ block $\times$ trial), with a median of 324 observations per participant. Including participant as a random intercept accounted for substantial between-participant variability in visual-aid reliance (intra-class correlation coefficient ICC ≈ 0.31), justifying the mixed-effects formulation. As shown in Figure 12, Ctrl-key use decreased steadily across blocks, reflecting a potential transition from visually guided search to memory-based recall. In addition, participants in the size-variant interfaces may have reduced their reliance on the visual aid more rapidly than in the Baseline, as suggested by faster spatial learning and recall in those interfaces.

4.3.6 Subjective Workload and Preferences. Subjective workload (NASA-TLX) scores were analyzed using ART [59]. Size variant interfaces were perceived as less demanding than the uniform Baseline, with lower reported mental and temporal effort, particularly for

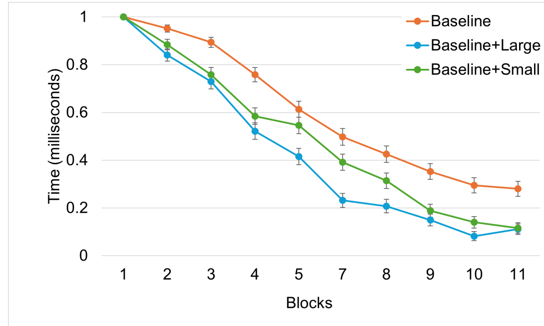


Figure 12: Proportion of Ctrl-key (visual aid) use across blocks for Study 2 interfaces. Error bars show standard error.

the Baseline+Large condition (Table 9). Participant preferences also strongly favoured Baseline+Large for speed, memorability, comfort, and overall usability, aligning with the observed performance advantages of size-based landmarks (Table 10).

Table 9: Study 2 mean NASA–TLX scores (0 = low, 7 = high). Standard deviations in parentheses. Performance is reversed.

Subscale	Interface			F	p
	Baseline	Baseline +Large	Baseline +Small		
Mental Demand	5.39 (1.31)	3.65 (1.77)	4.30 (1.72)	8.61	< .001
Physical Demand	3.43 (2.04)	2.43 (1.75)	3.26 (1.86)	4.88	.012
Temporal Demand	3.83 (1.99)	2.52 (1.50)	3.17 (1.44)	5.32	.009
Performance (rev.)	4.61 (1.44)	5.91 (1.31)	5.26 (1.45)	12.25	< .001
Effort	4.91 (1.56)	3.35 (1.75)	4.13 (1.66)	8.99	< .001
Frustration	4.04 (1.87)	2.57 (1.90)	3.43 (2.04)	9.96	< .001

Table 10: Study 2 participant preferences across interfaces.

Measure	Baseline	Baseline+Large	Baseline+Small
Speed	0	20	3
Accuracy	0	18	5
Memorization	0	19	4
Comfort	0	19	4
Overall	0	19	4

5 Discussion

5.1 Explanations for Results

5.1.1 Size Variation and Learning Toward Memory-Based Selection. Across both studies, performance improved with practice, consistent with the power law of practice [39]. Size-variant interfaces showed their clearest advantages in later blocks, when participants had repeated exposure to the same target set and could increasingly rely on learned object–location associations rather than sustained visual search. This pattern aligns with established models of skill acquisition, in which interaction shifts from visually guided behaviour toward memory-based execution with practice [19]. In Study 2, these benefits persisted in blind blocks, where labels could not be revealed, indicating that size variation supported durable spatial representations beyond momentary visual guidance. Decreasing reliance on the `Ctrl` visual aid further supports a transition toward memory-supported behaviour as practice accumulated.

5.1.2 Perceptual and Semantic Contributions. Comparing semantic and abstract interfaces clarifies the role of perceptual versus semantic cues. In Study 1, semantic icon meaning complemented size variation, supporting faster recall. In Study 2, where semantic information was removed, size-based landmarks alone continued to yield reliable benefits. This convergence demonstrates that perceptual distinctiveness can be useful to support spatial learning, consistent with theories of preattentive processing and visual salience [50, 60]. At the same time, semantic cues appear to strengthen spatial organization when available, likely by providing additional top-down associations [12, 53]. The findings, therefore, suggest that size variation can operate as a foundational perceptual cue that integrates naturally with higher-level semantic structure.

5.1.3 Effectiveness of Large Icons Landmarks. Across all analyses, enlarged icons produced the most consistent and robust benefits. Large icons served as stable global anchors that improved recall accuracy, reduced completion time, and reduced dependence on visual confirmation. Smaller icons offered more localized benefits, particularly in dense regions, but were less reliable overall. Interfaces combining multiple size variations showed reduced efficiency, likely due to competing salience and weakened visual hierarchy. This pattern is consistent with principles from graphical perception and information visualization, which caution against excessive heterogeneity [10, 13]. Spatial analyses further showed that size-based landmarks were most beneficial in central interface regions, where natural geometric anchors such as edges and corners were absent. In contrast, targets near boundaries were already efficient to locate, indicating that artificial landmarks primarily compensate for anchor-poor regions rather than uniformly improving all locations.

5.1.4 Landmarks as Spatial Anchors. Spatial position analyses indicate that landmarks shape how participants organize command locations. Enlarged items provided stable reference points, reducing the disadvantage of middle-region targets relative to natural boundary anchors (edges and corners), suggesting that intrinsic landmarks can compensate for anchor-poor regions in dense grids. The persistence of performance advantages under memory-only conditions in Study 2 supports the interpretation that size-based landmarks contributed to the formation of spatial representations, rather than guiding attention moment-by-moment, instead structuring retrieval over time [57].

5.2 Implications for Interface Design

Our findings suggest that *intrinsic* icon size variation can serve as a lightweight landmarking mechanism in command-dense, spatially stable GUIs. The following implications translate the observed performance patterns into practical design guidance.

5.2.1 Introduce a small number of enlarged icons as spatial anchors, particularly in central regions. Across both studies, enlarged items provided the most consistent recall advantages, with the strongest benefits observed for targets located away from natural boundaries. This suggests that placing a limited number of larger icons in central regions can compensate for the absence of edges and corners, helping users organize spatial memory around stable reference points.

5.2.2 Use reduced-size icons sparingly to support local differentiation, not global structure. Smaller icons offered more localized benefits and were less reliable than enlarged landmarks. Their use is best suited for breaking symmetry or disambiguating dense clusters, rather than serving as primary anchors. Overuse or excessive size variation risks perceptual fragmentation, consistent with established guidance on visual encoding and grouping [10, 13].

5.2.3 Maintain stable landmark placement across interaction sessions. Size-based landmarks supported learning primarily through repeated exposure and later-block performance. To preserve these benefits, landmark positions should remain stable across sessions and interface updates. Frequent relocation of anchors would require relearning and undermine the formation of durable object–location associations [14].

5.2.4 Align size landmarks with meaningful structure when semantics are available. When icons carried semantic meaning (Study 1), size variation complemented recognition-based cues; when semantics were removed (Study 2), size alone still supported recall. In practical interfaces, aligning enlarged landmarks with stable conceptual groupings can provide multiple retrieval pathways—by meaning and by location—without introducing additional visual elements.

These principles are consistent with prior work on spatial memory in interfaces [9, 14] and extend naturally to emerging contexts such as wearable and immersive systems, where stable spatial cues are critical for efficient recall-driven interaction.

6 Limitations and Future Work

Our studies were conducted in controlled settings using repeated selection tasks, which supports internal validity but may not capture the richer visual context, multitasking, and evolving command sets of real applications. Future work should test intrinsic size landmarks in task-oriented software (e.g., editing, visualization, development tools) and in interfaces where commands change over time.

Although blind blocks provided a strict test of memory-based selection, only two such blocks were included to avoid repeatedly interrupting normal interaction strategies. Future work could explore denser or delayed blind testing to examine longer-term retention and interference effects.

We focused on within-session learning. Although blind-block performance indicates that participants formed usable spatial representations, longer-term retention and transfer across days or devices remain open. Longitudinal designs could assess persistence, relearning costs, and cross-device generalization.

Finally, icon size is coupled with low-level salience. While blind blocks reduce reliance on visual confirmation, future work can further separate salience from landmarking by controlling luminance/area or comparing size against other intrinsic cues (e.g., outline weight). Eye-tracking could also directly characterize how attention shifts from visual search toward memory-supported behaviour during learning.

7 Conclusion

Across two controlled studies, we explored icon-size variation as an intrinsic landmarking mechanism to support spatial learning and memory-driven interaction in dense command interfaces. Results showed that enlarged icons consistently improved recall speed and reduced errors for both semantic and abstract icons, with benefits persisting under memory-only conditions. These findings show that perceptual size distinctiveness alone can support spatial memory development of commands, with larger icons providing a better memory aid than smaller ones, particularly in central regions lacking natural boundaries. We demonstrate that icon size variation can be a lightweight and scalable design strategy for facilitating memory-based interaction in command-dense computer interfaces.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC): RGPIN-2023-04393. We also thank the anonymous reviewers for their valuable suggestions.

References

[1] Sayeem Md. Abdullah and Md. Sami Uddin. 2024. Effects of Increasing Command Capacity of Spatial Memory Menus in Tablets. In *Companion Proceedings of the 2024 Conference on Interactive Surfaces and Spaces* (Vancouver, BC, Canada) *(ISS Companion '24)*. Association for Computing Machinery, New York, NY, USA, 69–72. doi:10.1145/3696762.3698055

[2] Sayeem Md Abdullah and Md. Sami Uddin. 2024. Revisiting FastTap: Effects of Increasing Command Capacity of Spatial Memory Menus in Tablets. In *2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, Yorktown Heights, NY, USA, 0683–0689. doi:10.1109/UEMCON62879.2024.10754660

[3] Sayeem Md Abdullah and Md. Sami Uddin. 2025. Effects of Artificial Landmarks on Spatial Learning in FastTap Menus with Large Command Sets. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 212, 7 pages. doi:10.1145/3706599.3720030

[4] Sayeem Md Abdullah and Md. Sami Uddin. 2025. Enhancing Spatial Learning of Large Command Sets in Two FastTap Menus with Artificial Landmarks. In *Proceedings of the Conference on Graphics Interface (GI '25)* (New York, NY, USA). Association for Computing Machinery, Kelowna, BC, Canada, 1–12.

[5] Jason Alexander, Andy Cockburn, Stephen Fitchett, Carl Gutwin, and Saul Greenberg. 2009. Revisiting read wear: analysis, design, and evaluation of a footprints scrollbar. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1665–1674. doi:10.1145/1518701.1518957

[6] Wendy S. Ark, D. Christopher Dryer, Ted Selker, and Shumin Zhai. 1998. Landmarks to Aid Navigation in a Graphical User Interface. In *Proceedings of the Workshop on Personalized and Social Navigation in Information Space*. IFIP WG 13.2, Swedish Institute of Computer Science (SICS), Stockholm, Sweden, 16–17. https://www.researchgate.net/publication/235328897_Landmarks_to_aid_navigation_in_a_graphical_user_interface SICS Technical Report 98:01.

[7] Tahsin Aziz and Md. Sami Uddin. 2026. Size as an Intrinsic Landmark: How Icon Size Variation Supports Spatial Learning in Graphical Interfaces. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems* (Barcelona, Spain) *(CHI EA '26)*. Association for Computing Machinery, New York, NY, USA, Article 598, 5 pages. doi:10.1145/3772363.3799081

[8] Gilles Bailly, Daniel Duarte, Antti Oulasvirta, and Luis A. Leiva. 2025. Modeling how menu search strategies develop with experience. *International Journal of Human-Computer Studies* 205 (2025), 103615. doi:10.1016/j.ijhcs.2025.103615

[9] Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2016. Visual Menu Techniques. *ACM Comput. Surv.* 49, 4, Article 60 (Dec. 2016), 41 pages. doi:10.1145/3002171

[10] Jacques Bertin. 1983. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, Madison, WI. Translated by William J. Berg.

[11] Jack Callahan, Don Hopkins, Mark Weiser, and Ben Shneiderman. 1988. An empirical comparison of pie vs. linear menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Washington, D.C., USA) *(CHI '88)*. Association for Computing Machinery, New York, NY, USA, 95–100. doi:10.1145/57167.57182

[12] Febi Chajadi, Md. Sami Uddin, and Carl Gutwin. 2020. Effects of Visual Distinctiveness on Learning and Retrieval in Icon Toolbars. In *Proceedings of Graphics Interface 2020*. Canadian Human-Computer Communications Society, Edmonton, Canada, 93–100. doi:10.20380/GI2020.12

[13] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554. doi:10.1080/01621459.1984.10478080

[14] Andy Cockburn, Carl Gutwin, and Saul Greenberg. 2007. A predictive model of menu performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 627–636. doi:10.1145/1240624.1240723

[15] Jacob Cohen. 1973. Eta-Squared and Partial Eta-Squared in Fixed Factor ANOVA Designs. *Educational and Psychological Measurement* 33, 1 (1973), 107–112. doi:10.1177/001316447303300111

[16] Philip M. Corsi. 1972. *Human Memory and the Medial Temporal Region of the Brain*. Ph. D. Dissertation. McGill University, Montreal, Canada. Unpublished doctoral dissertation.

[17] Mary P. Czerwinski, Maarten van Dantzich, George G. Robertson, and Hunter G. Hoffman. 1999. The Contribution of Thumbnail Image, Mouse-Over Text and Spatial Location Memory to Web Page Retrieval in 3D. In *Proceedings of INTERACT '99: IFIP TC13 International Conference on Human-Computer Interaction*. IOS Press, Edinburgh, UK, 163–170.

[18] Brian D. Ehret. 2002. Learning where to look: location learning in graphical user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02)*. Association for Computing Machinery, New York, NY, USA, 211–218. doi:10.1145/503376.503414

[19] Paul M. Fitts and Michael I. Posner. 1967. *Human Performance*. Brooks/Cole Publishing Company, Belmont, CA.

[20] Krzysztof Z. Gajos, Mary Czerwinski, Desney S. Tan, and Daniel S. Weld. 2006. Exploring the design space for adaptive graphical user interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Venezia, Italy) *(AVI '06)*. Association for Computing Machinery, New York, NY, USA, 201–208. doi:10.1145/1133265.1133306

[21] BoYu Gao, HyungSeok Kim, Byungmoon Kim, and Jee-In Kim. 2018. Artificial Landmarks to Facilitate Spatial Learning and Recalling for Curved Visual Wall Layout in Virtual Reality. In *Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, Shanghai, China, 475–482. doi:10.1109/BigComp.2018.00076

[22] Carl Gutwin and Andy Cockburn. 2006. Improving list revisitation with ListMaps. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Venezia, Italy) *(AVI '06)*. Association for Computing Machinery, New York, NY, USA, 396–403. doi:10.1145/1133265.1133347

[23] Carl Gutwin, Andy Cockburn, Joey Scarr, Sylvain Malacria, and Scott C. Olson. 2014. Faster command selection on tablets with FastTap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2617–2626. doi:10.1145/2556288.2557136

[24] Christopher Healey and James Enns. 2012. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics* 18, 7 (July 2012), 1170–1188. doi:10.1109/TVCG.2011.127

[25] William E. Hick. 1952. On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology* 4, 1 (1952), 11–26. doi:10.1080/17470215208416600

[26] William C. Hill, James D. Hollan, Dave Wroblewski, and Tim McCandless. 1992. Edit wear and read wear. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 3–9. doi:10.1145/142750.142751

[27] Stephen C. Hirtle and John Jonides. 1985. Evidence of Hierarchies in Cognitive Maps. *Memory & Cognition* 13, 3 (1985), 208–217. doi:10.3758/BF03197683

[28] Icons8. 2018. Line Awesome Icon Set. https://icons8.com/line-awesome. Accessed: 2025-08-18.

[29] Kurt Koffka. 1935. *Principles of Gestalt Psychology*. Harcourt, Brace and Company, New York, NY. Reprinted by Routledge, 1999.

[30] Gordon Kurtenbach and William Buxton. 1994. User learning and performance with marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 258–264. doi:10.1145/191666.191759

[31] Benjamin Lafreniere, Carl Gutwin, Andy Cockburn, and Tovi Grossman. 2016. Faster Command Selection on Touchscreen Watches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4663–4674. doi:10.1145/2858036.2858166

[32] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. 2012. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics* 18, 9 (Sep. 2012), 1520–1536. doi:10.1109/TVCG.2011.279

[33] J. P. Lewis, Ruth Rosenholtz, Nickson Fong, and Ulrich Neumann. 2004. VisualIDs: automatic distinctive icons for desktop interfaces. *ACM Trans. Graph.* 23, 3 (Aug. 2004), 416–423. doi:10.1145/1015706.1015739

[34] Jiazhou Liu, Kadek Ananta Satriadi, Barrett Ens, and Tim Dwyer. 2024. Investigating the Effects of Physical Landmarks on Spatial Memory for Information Visualisation in Augmented Reality. In *Proceedings of the 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Seattle, WA, USA, 289–298. doi:10.1109/ISMAR62088.2024.00043

[35] Kevin Lynch. 1960. *The Image of the City*. MIT Press, Cambridge, MA.

[36] Aristides Mairena, Md. Sami Uddin, and Carl Gutwin. 2022. Accidental Landmarks: How Showing (and Removing) Emphasis in a 2D Visualization Affected Retrieval and Revisitation. In *Proceedings of Graphics Interface 2022*. Canadian Human-Computer Communications Society, Toronto, Canada, 53–62. doi:10.20380/GI2022.07

[37] Sylvain Malacria, Gilles Bailly, Joel Harrison, Andy Cockburn, and Carl Gutwin. 2013. Promoting Hotkey use through rehearsal with ExposeHK. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 573–582. doi:10.1145/2470654.2470735

[38] Daniel R. Montello. 2001. Spatial Cognition. In *International Encyclopedia of the Social and Behavioral Sciences*, Neil J. Smelser and Paul B. Baltes (Eds.). Elsevier, Oxford, UK, 14771–14775. doi:10.1016/B0-08-043076-7/02492-X

[39] Allen Newell and Paul S. Rosenbloom. 1981. Mechanisms of Skill Acquisition and the Law of Practice. In *Cognitive Skills and Their Acquisition*, John R. Anderson (Ed.). Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1–55.

[40] Daniel L. Odell, Richard C. Davis, Andrew Smith, and Paul K. Wright. 2004. Toolglasses, marking menus, and hotkeys: a comparison of one and two-handed command selection techniques. In *Proceedings of Graphics Interface 2004* (London, Ontario, Canada) *(GI '04)*. Canadian Human-Computer Communications Society, Waterloo, CAN, 17–24.

[41] Amy Pavel, Dan B. Goldman, Björn Hartmann, and Maneesh Agrawala. 2015. SceneSkim: Searching and Browsing Movies Using Synchronized Captions, Scripts and Plot Summaries. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 181–190. doi:10.1145/2807442.2807502

[42] Amy Pavel, Colorado Reed, Björn Hartmann, and Maneesh Agrawala. 2014. Video digests: a browsable, skimmable format for informational lecture videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 573–582. doi:10.1145/2642918.2647400

[43] Albert Postma and Edward H. F. De Haan. 1996. What Was Where? Memory for Object Locations. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology* 49, 1 (Feb. 1996), 178–199. doi:10.1080/713755605

[44] Clark C. Presson. 1987. The Development of Landmarks in Spatial Memory: The Role of Differential Experience. *Journal of Experimental Child Psychology* 44, 3 (1987), 317–334. doi:10.1016/0022-0965(87)90037-3

[45] Joey Scarr, Andy Cockburn, Carl Gutwin, and Andrea Bunt. 2012. Improving command selection with CommandMaps. 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, 257–266. doi:10.1145/2207676.2207713

[46] Joey Scarr, Andy Cockburn, Carl Gutwin, and Philip Quinn. 2011. Dips and ceilings: understanding and supporting transitions to expertise in user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2741–2750. doi:10.1145/1978942.1979348

[47] Katherine Schramm, Carl Gutwin, and Andy Cockburn. 2016. Supporting Transitions to Expertise in Hidden Toolbars. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4687–4698. doi:10.1145/2858036.2858412

[48] Alexander W. Siegel and Sheldon H. White. 1975. The Development of Spatial Representations of Large-Scale Environments. In *Advances in Child Development and Behavior*, Hayne W. Reese (Ed.). Vol. 10. Academic Press, New York, NY, 9–55. doi:10.1016/S0065-2407(08)60007-5

[49] Perry W. Thorndyke and Sarah E. Goldin. 1983. Spatial Learning and Reasoning Skill. In *Spatial Orientation: Theory, Research, and Application*, Herbert L. Pick and Linda P. Acredolo (Eds.). Springer US, Boston, MA, 195–217. doi:10.1007/978-1-4615-9325-6_9

[50] Anne M. Treisman and Garry Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (1980), 97–136. doi:10.1016/0010-0285(80)90005-5

[51] Md. Sami Uddin. 2022. *Use of Landmarks to Improve Spatial Learning and Revisitation in Computer Interfaces*. Ph. D. Dissertation. University of Saskatchewan.

[52] Md. Sami Uddin and Carl Gutwin. 2016. Rapid Command Selection on Multi-Touch Tablets with Single-Handed HandMark Menus. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces* (Niagara Falls, Ontario, Canada) *(ISS '16)*. Association for Computing Machinery, New York, NY, USA, 205–214. doi:10.1145/2992154.2992172

[53] Md. Sami Uddin and Carl Gutwin. 2021. The Image of the Interface: How People Use Landmarks to Develop Spatial Memory of Commands in Graphical Interfaces. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 515, 17 pages. doi:10.1145/3411764.3445050

[54] Md. Sami Uddin, Carl Gutwin, and Andy Cockburn. 2017. The Effects of Artificial Landmarks on Learning and Performance in Spatial-Memory Interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3843–3855. doi:10.1145/3025453.3025497

[55] Md. Sami Uddin, Carl Gutwin, and Alix Goguey. 2017. Using artificial landmarks to improve revisitation performance and spatial learning in linear control widgets. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 48–57. doi:10.1145/3131277.3132184

[56] Md. Sami Uddin, Carl Gutwin, and Benjamin Lafreniere. 2016. HandMark Menus: Rapid Command Selection and Large Command Sets on Multi-Touch Displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5836–5848. doi:10.1145/2858036.2858211

[57] Norman G. Vinson. 1999. Design guidelines for landmarks to support navigation in virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 278–285. doi:10.1145/302979.303062

[58] Colin Ware. 2013. *Information Visualization: Perception for Design* (3rd ed.). Morgan Kaufmann, Waltham, MA.

[59] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963

[60] Jeremy M. Wolfe. 1994. Guided Search 2.0: A Revised Model of Visual Search. *Psychonomic Bulletin & Review* 1, 2 (1994), 202–238. doi:10.3758/BF03200774

[61] Jeremy M. Wolfe. 1994. Visual search in continuous, naturalistic stimuli. *Vision Research* 34, 9 (1994), 1187–1195. doi:10.1016/0042-6989(94)90300-X