Effects of Visual Distinctiveness on Learning and Retrieval in Icon Toolbars

A thesis submitted to the
College of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Febi Chajadi

Permission to Use

Disclaimer

Abstract

Learnability is important in graphical interfaces because it supports the user's transition to expertise. One aspect of GUI learnability is the degree to which the icons in toolbars and ribbons are identifiable and memorable – but current "flat" and "subtle" designs that promote strong visual consistency could hinder learning by reducing visual distinctiveness within a set of icons. There is little known, however, about the effects of visual distinctiveness of icons on selection performance and learnability. To address this gap, we carried out two studies using several icon sets with different degrees of visual distinctiveness, and compared how quickly people could learn and retrieve the icons. Our first study found no evidence that increasing colour or shape distinctiveness improved learning, but found that icons with concrete imagery were easier to learn. Our second study found similar results: there was no effect of increasing either colour or shape distinctiveness, but there was again a clear improvement for icons with recognizable imagery. Our results show that visual characteristics appear to affect UI learnability much less than the meaning of the icons' representations.

Acknowledgements

Words cannot express my gratitude to my supervisor Carl Gutwin for his bottomless patience and generous feedback throughout my time at the University of Saskatchewan. His guidance has made me a better researcher, writer, and critical thinker.

Many thanks to my thesis committee whose suggestions and comments were invaluable.

I'd like to acknowledge the staff of the CS department for always being there to make our lives a little less stressful.

I would be remiss in not mentioning my friends at the HCI Lab whose support you can always count on during rocky times.

This endeavour would not have been possible without my mom and sister who were there every step of the way to prop me up when I couldn't.

Contributions of Authors

The work presented in this thesis is published at Graphics Interface 2020 with co-authors Drs. Carl Gutwin and Sami Uddin. Dr. Gutwin's contributions to the project include guidance in the research direction, information and writing related to domain-specific knowledge and background, advice on experimental design, and editing of the conference manuscript. Dr. Uddin's contributions include advice on experimental design and writing of related work in the conference manuscript.

This thesis is dedicated to my dad whose dream was to see his girls graduate college. I may have overdone it a little.

CONTENTS

List of Tables

List of Figures

1 INTRODUCTION

Learnability—that is, “a novice user’s experience of the initial part of the learning curve” [124] which, within exceedingly learnable systems, would allow “users to reach a reasonable level of usage proficiency within a short time” [66]—is important in graphical user interfaces because it is an important part of a user’s transition from novice to expert. Many kinds of learning can occur with an interface, but for WIMP interfaces (systems with windows, icons, menus, and pointers), one main way that users improve their performance is by learning the commands associated with icons in toolbars and ribbons, and where those icons are located. Therefore, a goal in the visual design of icons is to help the user remember the icon and the underlying command. However, other goals in icon design may interfere with an icon’s ability to support this learning process. One of these goals is the desire for visual consistency and cohesiveness-—the idea that all of the icons in an interface should repeat the same visual variables (such as colour, contrast, weight, shape, angle, and size) to tie together the visual elements of the interface and give the system a recognizable style.

Figure 1.1: This figure has been removed due to copyright restrictions. It was an example of “visually cohesive icon design”. Visual variables such as colour, contrast, weight, and size were repeated across the entire set. Original source: https://symbolset.com/icons/standard

For example, Figures 1.1 and 1.2 show icons presented as good examples of visual consistency in icon design. These icons also illustrate a second design goal that is common in many commercial systems-—subtle and “flat” icon design, in which icons are monochrome and have relatively low contrast.

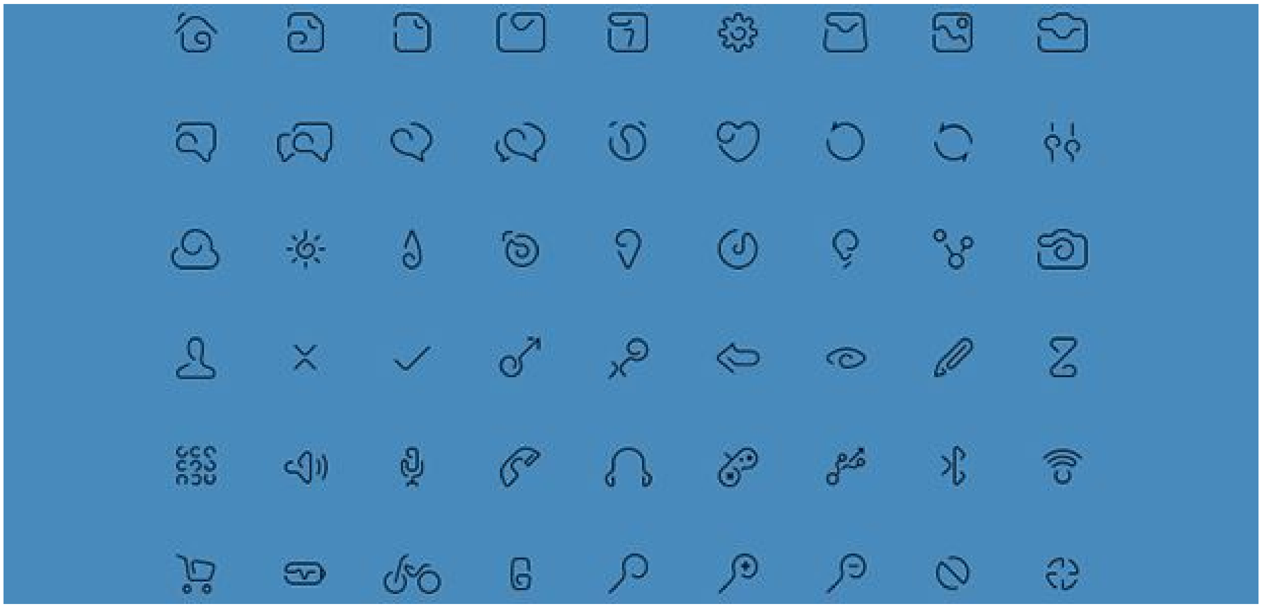


Figure 1.2: Example web post of “minimalist icon design” [166]. Visual variables such as colour, contrast, weight, and size are repeated across the entire set. From Outline Icons [41].

1.1 Problem and Motivation

Although visually consistent icons are popular, the similarity across several visual variables also reduces distinctiveness, which could hinder visibility and learnability (for example, if all of a UI’s icons were identical grey rectangles, they would presumably be difficult to remember). More generally, it seems likely that icon learnability could be affected by the visual attributes of the icons. This issue has been raised by some users who have noticed the potential problems of “flat” icon design: for example, forum posts often complain that icons are too similar [87, 139, 177, 45]. Recently, Microsoft has moved away from flat icons to more colourful and non-uniform imagery [92, 191] (Figure 1.3) reminiscent of previously popular graphical design trends (as shown in Figure 1.4).

Figure 1.3: This figure has been removed due to copyright restrictions. It was a diagram of Microsoft Mail icons over the past 20 years. Original source: https://medium.com/microsoft-design/iconic-icons-designing-the-world-of-windows-5e70e25e5416

In this thesis, we address the lack of research into the effects of visual variables (differentiable properties in graphical objects such as colour, position, size, orientation etc.) on the learnability of icons. Designers need to know whether the visual properties of icons can affect learnability. Existing research on the effects of visual variables in information

Figure 1.4: 1995, SVGA: colourful and distinguishable icons. Used with permission from Microsoft [110]

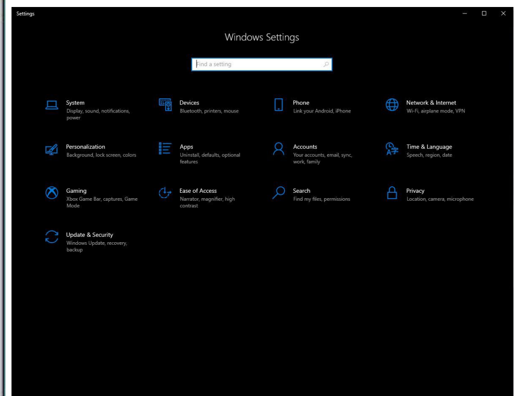


Figure 1.5: 2019, 4K and millions of colours: flat monochrome icons. Used with permission from Microsoft [110]

visualization and attention suggests an increase in both noticeability and learnability when graphical objects employ variations in colour and shape [179, 161, 16]. Certain visual properties (e.g., suggested contrast ratios between text and background [10]) have also been investigated for their usability. However, there is little known about the effects of visual variables on learning. Furthermore, additional consideration needs to be made when investigating shape distinctiveness as distinctiveness does not only encompass the physical properties of an icon; an icon's shape can carry meaning which could be useful for learning as people can quickly find recognizable shapes in an interface [14, 53]. However, it is not always possible to convey accurate meaning for abstract concepts in icons (e.g., icons for abstract or complex commands such as "Analyse" represented as a formula sign or barcodes as "Encoding") which may cause confusion and impede learning. As such, unlike icons in information visualization, investigating shape distinctiveness in graphical user interface icons involves investigating the effect of meaning in icon learnability. The role of visual variables in icon learnability needs to be addressed to assess the effect of modern icon design guidelines on learning.

1.2 Solution

To examine the effects of shape and colour as factors on the learnability of icons, we carried out two user studies to compare how quickly people could learn and select icons with varying degrees of visual distinctiveness. Our first study tested the effects of shape distinctiveness by comparing icons that had more-similar shapes to icons with less-similar shapes, and icons that conveyed no meaning (abstract) to icons whose shape showed an object or symbol (concrete). We also tested the effects of colour by comparing monochrome icons to icons of different colours. Participants were asked to find and select target commands from a toolbar with 60 icons, repeated over five blocks. We gauged learnability through completion time, hover amounts, subjective rankings, and participant comments.

As will be demonstrated later in our first study, prestudy, and second study, in the absence of meaning, participants would employ mnemonics to facilitate learning icons that were otherwise meaningless or of a context unfamiliar to them. Consequently, we identified a gap that was not addressed in our previous studies regarding the familiarity of an icon's imagery on its learnability. Thus, our second study tested the effects of three factors in a series of planned comparisons. We assessed shape distinctiveness (icons were either identical squares or concrete images), colour (squares could be either monochrome or coloured) and a new factor, familiarity (concrete icons were either unfamiliar shapes or familiar images). Completion time, hover amounts, incorrect selections, subjective rankings and participant comments were again used to measure the learnability of these factors in our hypotheses. https://www.overleaf.com/project/62cee07f021579457161560a

1.3 Steps to the Solution

Evaluating the effects of visual variables on learnability in icons comprised multiple steps:

- **Identify visual variables to examine learnability**

  In order to examine the learnability of icons, we first had to identify the visual variables that affected noticeability and learnability. Colour has been extensively studied in psychology and information visualization as a powerful tool to attract attention [201, 179, 178, 89]—–an important precursor to the formation of memory—– but it is uncertain how it affects learnability in icons. We theorize that colour (especially those found to attract attention) would positively affect the learnability of icons by allowing users to search and filter by colour in a field of icons when prompted. Similarly, differences in shape have been suggested as an effective tool to increase noticeability and learnability [201]. We posit that noticeable differences in shapes may act as memory hooks or visual landmarks that increase learnability. From design guidelines [19, 117] to semiotics research [132, 141], meaning has been shown to affect icon usage and learnability. Results suggest that icons that represent their underlying command as closely as possible tend to be easier to navigate, remembered and favoured by users compared to abstract icons that do not represent their underlying commands.

- **Implement system to evaluate the effect of colour in icons with differing levels of shape distinctiveness**

  Following the identification of suitable visual variables (colour and shape), we created five interfaces to facilitate the assessment of the aforementioned factors on learnability. Colour was evaluated in icons with varying levels of shape distinctiveness. Icons with no shape distinctiveness consisted of recognizable shapes that were unified by a border and grey background for visual consistency and ensuring shapes that were more unique would not affect results. Icons with shape distinctiveness were either square or circular shaped. The square or circular shaped icons were differentiated by lines drawn within their shape boundaries. An additional icon set of colourful varied shapes and colour was also evaluated to ascertain benefits of combining meaning and colour without any unifying theme.

- **Determine number of colours to use**

  Prior to the second study, a prestudy was conducted to determine whether to use a four, eight or twelve colour icon set to study the effects of colour as the only visual variable. Empirical evidence, participant comments and general observations were used to decide the number of colours to use.

- **Implement icon sets of varying colour or shape distinctiveness**

  While the first study determined that increased colour distinctiveness did not improve learnability in icons with and without shape distinctiveness, we were unable to determine whether or not it was due to colour under-performing or the out-sized effects of the interfaces' other visual variables. For example, in the coloured interface without meaning, each icon also possessed texture (various lines within the borders of the icon) as an additional visual variable to colour. We were also unable to determine if meaning in the interfaces with meaning was an overriding factor in learnability. To evaluate the effect of our factors without possible additive (or negative) effects of combining with other visual variables, we created four icon sets to see how colour, shape, and meaning performs.

1.4 Evaluation

Variances in meaning (none, contextual, familiar), shape distinctiveness (none, medium, high) and color distinctiveness (none, medium) were evaluated across two user studies to determine the effects of the aforementioned on learnability. The first study investigated the effects of colour by comparing monochrome icons to icons of different colours. Shape distinctiveness was also evaluated by comparing abstract icons to concrete icons. The second study evaluated colour, shapes and meaning as individual factors. Learnability was assessed through completion time, incorrect selections and hover amounts. Subjective responses were also collected to determine icon sets that were easiest, hardest, and preferred. NASA TLX style questionnaires were collected to gauge perceived performance and perceived mental, physical, and temporal load.

Our first study provided the following findings:

- Icons with concrete imagery were learned much faster than abstract icons;
- Adding colour to either the concrete or the abstract icon set did not lead to improved learning;
- Varying both shape and colour did not improve learnability in concrete icons;

Our second study provided the following findings:

- The addition of colour to identical grey squares did not improve learning;
- Unfamiliar shapes (Kanji and Mandarin characters for users with no experience in these languages) were much harder to learn than familiar shapes (everyday objects);
- There was no difference in learning between the Kanji characters and the grey squares, even though the characters were far more differentiable in terms of shape.

1.5 Contribution

The primary contribution of this thesis is new information about how visual distinctiveness affects icon learning and retrieval. Surprisingly, the low visual distinctiveness of “flat” and subtle icon designs does not appear to make them more difficult to find or remember. Instead, having a concrete visual representation in the icon was shown to be extremely valuable for learning the icons. Based on our participants’ comments, we suggest that this property better allows users to create a “memory hook” for the association between the icon and the command. Minor contributions in this thesis include new information about the lack of learnability improvements in icons without concrete meaning. Despite the addition of visual variables such as colour, or shape, learnability in icons without concrete meaning were not significantly different from the learnability of an icon set with no visually discernible differences. Furthermore, colour did not prove to be an additive improvement in learning in icons with or without concrete meaning. Our findings also contribute to a better understanding of how visual variables affect the process of learning icon locations and provide a clear suggestion to designers that concrete images are likely to be more important than the distinctiveness of visual variables.

1.6 Thesis Outline

Chapter Two presents a review of related research. In the first part of our review, we examine the biological processes behind the perception of visual imagery. First, we discuss light as the source of visual stimuli in the surrounding environment. Second, the human receptor system is discussed. Finally, we discuss the visual cortex.

In the second part of our review, we evaluate the cognitive processes that enable perception of visual imagery. First, visual attention and search are discussed. Second, we discuss the design of visual icons and literature on the effects of visual variables in multimedia and information visualization that determined the selection of visual variables to evaluate in our user studies. Finally, spatial memory and its effects on increasing recall efficiency as well as the development of expertise are discussed.

Chapter Three describes the web-based system that was developed for the two user studies to measure learnability in icon sets. The system presents a ribbon-like menu interface of sixty icons in three rows that participants will have to select prompted icons from. In this chapter, we provide an overview of the framework we derived from reviewing existing research in naming our factors and its levels used throughout both user studies. The tasks and stimuli participants were exposed to within our two user studies are discussed. Questionnaires and subjective measures are included as well as the equipment both user studies were conducted on.

Chapter Four presents the first of two user studies. The first study evaluated the effect of colour on learnability in icons of various shape distinctiveness. Our four hypotheses for study 1 are introduced. A description of the interfaces used in the study is given as well as demographics and recruitment methods. Empirical results and subjective measures are presented and analyzed.

Chapter Five presents a prestudy to Study 2. The prestudy was conducted to determine the number of colours to use in our last user study as well as offer preliminary insights into shapes and meaning as the only available visual variables. Interface description is given as well as demographics and recruitment methods. Empirical results and subjective measures are presented and analysed.

Chapter Six presents the last of the user studies. We evaluated the effects of colour, meaning or shape as the only visual variable available in an interface. To examine colour, we used an icon set of identical squares, each icon coloured one of four colours. To examine shape distinctiveness, we used an icon set composed of four-stroke Chinese characters. To examine meaning, we used an icon set of recognizable real-world objects however, in a visually consistent manner by superimposing the icons on grey circular backgrounds. We also included an additional condition of identical grey squares—an interface without any visually discernible distinctiveness—to measure medium colour distinctiveness, high shape distinctiveness and meaning against. Our four hypotheses for study 2 are discussed. Procedure description is given as well as demographics and recruitment methods. Empirical results and subjective measures are presented and analysed.

Chapter Seven provides a discussion of the results from the previous three chapters and summarises the implications of our findings. Limitations of the studies, as well as potential future work are addressed. A summary of our findings and contributions are included and concludes this thesis.

2 Related Work

Learning and retrieving icons in toolbars involves several perceptual and cognitive processes. The task of learning icons in toolbars begins with the processing of light stimuli that make up the icons by the human visual system. Without the ability to see and to discern just what we're seeing, we cannot begin to familiarize ourselves with the ribbon toolbars we encounter. Thus this research is built on three foundational areas: the human visual system, design of visual icons, and spatial memory and the development of expertise to investigate whether visual variables affect learning.

2.1 Background in Visual Perception, Visual Attention, and Visual Search

Perception of visual imagery is a complex process relying on cognitive and behavioural processes to interpret sensory information. Humans' sense of sight constructs a representation of the surrounding environment from visible light that travels through the visual system as electrochemical signals [194]. This signal terminates at the visual cortex, where the stimuli are processed into an integrated whole [80].

This passive receipt of stimuli, however, is shaped by learning, memory, and attention [64, 17]. For example, a bottleneck exists in the brain for processing visual data: only 1% of incoming visual stimuli can be processed per second [210, 62], and as such, the role of attention in memory formation is crucial as attention focuses people's limited cognitive resources [9]. Furthermore, the quality of memory relies on the amount of attention directed at a stimulus [51]. Certain properties of visual images such as movement or colour have also been found to attract attention [207]. The following sections will briefly examine the processing of stimuli through the human visual system and how we attend to stimuli.

2.1.1 Visual Perception

Visual stimuli are composed of waves of electromagnetic energy. Wavelengths (the distance between oscillations) and amplitudes (the height of oscillations) in a wave determine the hue and brightness of a given stimulus [195]. The human eye can perceive wavelengths of 380 to 750 nanometers [168]. Furthermore, stimuli can also consist of wavelengths diluted by achromatic light, which are gray or white that do not register on the visible electromagnetic spectrum [195]. As the human eye is trichromatic, the hues in waves are typically expressed as a combination of the three primary colours (red, green, and blue) [206].

These stimuli are introduced to the human visual system pipeline through the receptor system which serves a crucial part in human visual performance by transducing electromagnetic energy into electrochemical neural energy that can be transported by the optic nerve to the brain in the form of electrical impulses [121]. The property of a stimulus depends upon where in the retina the light contacts. The retina is populated by rod and cone receptor cells that convert light into electrical signals [138]. Cone cells, which perceive colour and high degrees of detail in visual images, are further divided into three types: S, M, and L [142]. L, or long, cone cells are the most commonly found cells in the retina. They respond to light within the range of 500-700 nm, peaking between 564-580 nm depending on the amount of opsin within the cells a given individual possesses. The light produced in this band of wavelengths is typically perceived as reddish in hue (Figure 2.1). Medium (M) cells, which constitute approximately a third of the cone cells in the retina, respond to wavelengths in the 450-630 nm range with a peak between 534 to 555 nm. These wavelengths are commonly perceived as green in hue. S, or short, cone cells are the least commonly found cells in the retina, making up approximately 2% of all cone cells. They respond to wavelengths within the 400-500 nm range, peaking between 420 and 440 nm and visually present as blue in hue.

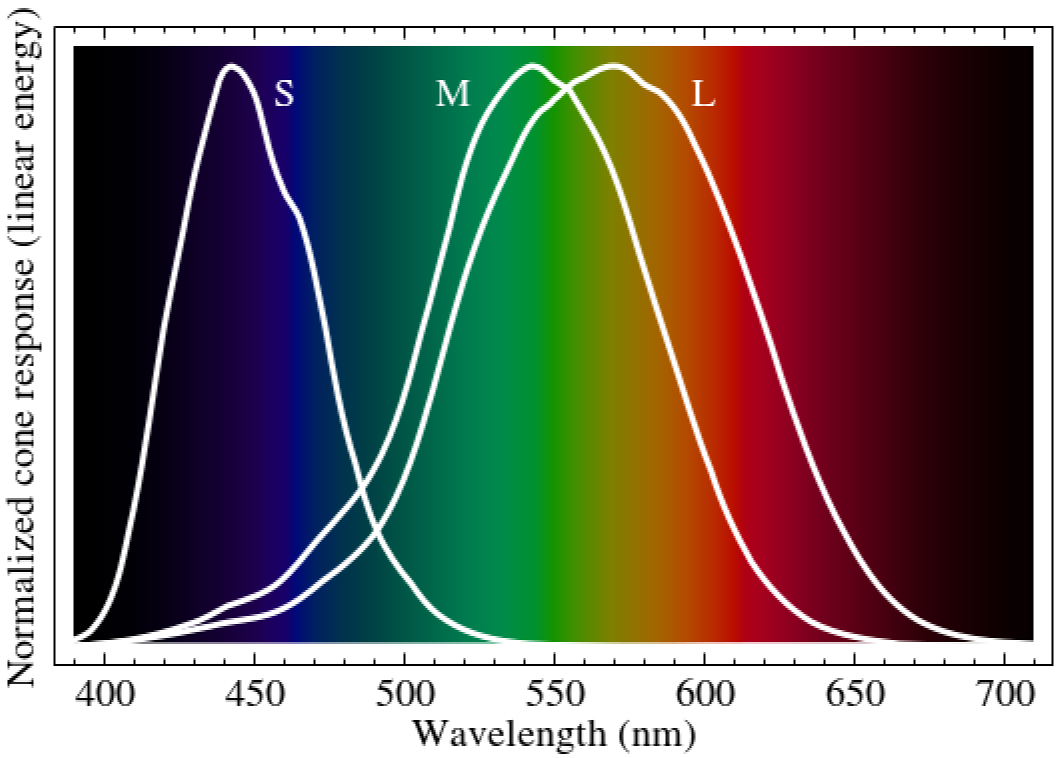


Figure 2.1: The wavelength ranges of short, medium, and long cone cells mapped against the spectrum of visible light. From "Normalized response spectra of human cones, to monochromatic spectral stimuli, with wavelength given in nanometers" [144].

The fovea, which inhabits a small area in the center of the retina, is primarily composed of cone cells. These cells excel in perceiving a high degree of detail and differentiating wavelengths of light (colour) [88, 153]. While it is less sensitive to light stimulation, it operates in a diminished manner in dim illumination. The periphery—that is, the area outside of the fovea—is populated by both rod cells and decreasing concentrations of cone cells the farther away from the center of the retina. Rod cells have higher sensitivity than cone cells and require less light to function. While rod cells cannot discriminate colour [126] or fine detail, they can perceive brightness, shape, and size. These cells allow the human eye to "see" in low light conditions but are less adaptable [190]. When exposed to light stimulation, rod cells lose sensitivity and require time to regain it in darkness. In practice, while colours can still be perceived in the periphery, they desaturate as the stimulus moves away from fixation (center of gaze) [86].

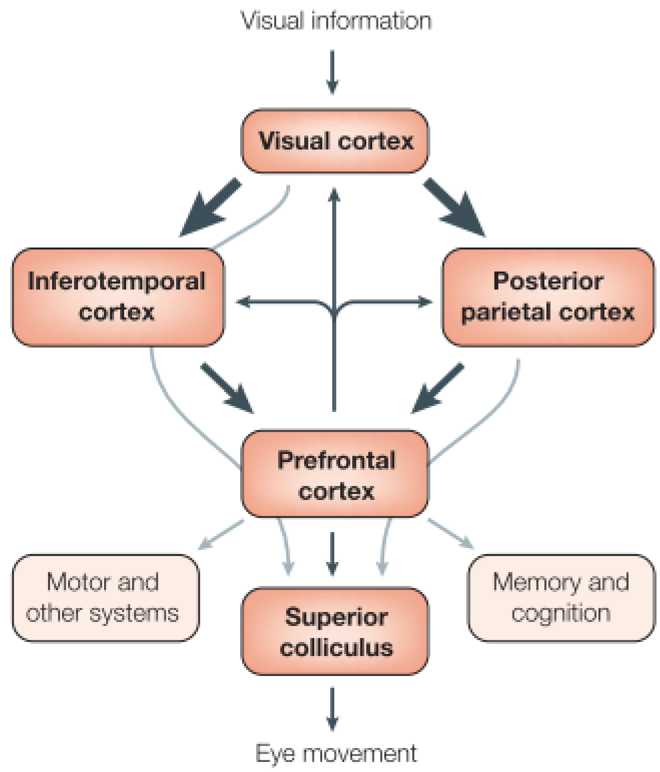


Figure 2.2: Flow chart of visual stimuli processing. From "Computational modeling of visual attention" [83]. This figure is being used in accordance with a University of Saskatchewan library licence agreement.

The lateral geniculate nucleus (LGN) and, to a lesser degree, the superior colliculus are the primary pathways for the electrical impulses created by the receptor system to reach the visual cortex [83]. The lateral geniculate nucleus consists of six layers of neurons [21]. The innermost two layers comprise magnocellular cells that transport input relating to motion, depth, and differences in brightness. The outermost four layers include parvocellular cells, which are integral to the perception of shape and colour as these neurons are sensitive to shape and colour. In between the six layers reside koniocellular cells. The purpose of the koniocellular layers in the LGN is not well understood; however, it has been established that these neurons receive input from short cone cells and, as such, may play a role in colour perception [172].

Information in the visual cortex is processed along two parallel hierarchical streams—dorsal and ventral—that build upon the incremental analysis of stimuli as it passes proceeding stages in the visual processing pathway of the brain [63]. The dorsal stream receives input from the magnocellular layers of the LGN through the 4Ca sub-layer of the primary visual cortex (V1). The neurons of the 4Ca sublayer are sensitive to motion in orientated bars, moving edges [79] and have demonstrated some attunement to direction [127]. Following the encoding in V1, the dorsal pathway projects stimuli into the secondary visual cortex (V2), which, while exhibiting no discernible purpose in the further encoding of information of the stimulus, is important in forwarding stimuli to the middle temporal visual area (MT/V5) [46]. The MT region of the visual cortex is populated by cells that are sensitive to two-dimensional motion on both local and global scopes (e.g., discerning both a group of moving circles and the individual circles within) [109]. The medial superior temporal area also processes motion, but for more complex, three-dimensional stimuli [133]. The dorsal stream terminates in the posterior parietal complex, which is responsible for processing complicated imagery and spatial localization. For example, it processes the motion of objects as the viewer is in motion and steers attention to objects of interest in the visual field [134].

The ventral stream primarily receives input from the parvocellular cells in the LGN. Whereas the dorsal stream establishes the "where", the ventral stream parses the "what" of visual imagery. It begins in the 4Cb, 4A, 3D, 2A, and 3A sub-layers of the V1. In these sub-layers, neurons code stimuli as either colour blobs or shape interblobs. Blobs are sensitive to colour but not contrast or size [164]. Interblobs are selective towards orientation and can process features [79]. Following encoding in the V1 region of the visual cortex, stimuli are projected into the V2 region, where neurons discern the orientation of edges [189] and assign them to objects, forming an early contour-based object representation [213]. Little to no encoding is done in regards to colour. Stimuli are next projected into the visual area V4 of the visual cortex, wherein cells tuned to hue discern colours [38]. The cells of the V4 region incorporate responses from the V1 and V2 regions to determine angles and curvatures [131]. V4 cells are also directionally selective—that is, sensitive to motion features [54]. The inferotemporal cortex (the terminus of the ventral stream) recognizes and identifies objects.

While the ventral stream does not control attention, it is nonetheless consequential to the processes in the dorsal stream as lesions that inhibit the ventral stream have been shown to create blind spots. It is reasonable to conclude that both the dorsal and ventral streams are necessary to deploy attention to stimuli successfully. However, while dependent, the two streams do not interact directly. As it is bidirectionally connected to both the dorsal and ventral streams, the prefrontal cortex is responsible for communication between the two streams [112]. Understanding the human visual perception system is critical to a holistic understanding of the task of learning icons in toolbars. While we do not measure any of these factors in our work, understanding this helps us design our study system, and interpret and explain findings of visual search tasks.

2.1.2 Visual Attention

A considerable amount of literature has been published on attention. Much of the existing literature in perceptual psychology pays particular attention to two prominent theories: top-down and bottom-up. Earlier bodies of literature favoured approaches that conform to the naturally perceived impression of objects—that is, objects are perceived as a whole rather than a combination of various features [115, 120]. For example, top-down processing proposes that visual scenes are registered in their entirety before the perception is disassembled into individual components as necessary. Owing to mechanisms that incorporate active cognitive participation (i.e., seeking blue socks when finding the other blue sock in the sock drawer), top-down mechanisms can often be observed in visual searching tasks [37].

However, physiological evidence suggests that the visual cortex responds to selective visual features such as colour and motion in the preattentive stage [207]. The bottom-up model posits that features in a visual scene are registered first and automatically in attention. In practice, bottom-up mechanisms allow for the diversion of attention towards items of increased salience—that is, degrees of difference from other elements in the scene (e.g. moving object in a static scene or the red dot in Figure 2.3). However, without focused attention—that is, the intentional concentration on particular stimuli by viewers—features cannot integrate into the objects we perceive. To address this shortcoming, Treisman et al. [179] developed a framework that identified separable visual features in the preattentive stage and the role focused attention plays in distinguishing presented objects when more than one visual feature is needed to recognize objects. Treisman's theory allowed that a "master map" of features is generated in the preattentive stage for every feature that occurs within a scene. When attention is given to a feature within a scene, it focuses the features, which are then stored in "object files." An association between the "object files" and prior knowledge is formed when the object is familiar.

Visual attention can be biased towards spatial—that is, location in the scene—or non-spatial properties [56]. An important aspect of visual search, feature-based attention allows for observers to search for stimuli amongst distractors by seeking specific features [170]. Studies suggest an improvement of behavioural performance and detection across the visual field when feature-based attention is involved [197]. Whether certain features are more suited than others for feature-based attention search, however, has yet to be conclusively determined [25]. While location information plays an important role in separating signal from noise in feature-based attention [180, 94], the effects of some non-spatial information can be contentious [174]. Generally, studies support the benefit of non-spatial features in improving the ingress of information into the brain [99]. It is important to understand visual attention because the task of learning icons in a toolbar involves processes such as top-down processing when seeking target icons. Furthermore, it is imperative to understand the effects that icons' visual properties may impart to avoid unintentional biases as we design our study interfaces.

2.1.3 Visual Search

Visual search—an integral part of collecting human visual performance metrics when measuring attention [198]—is the perceptual task of searching for a target among a field of distractors (non-target items). A common method for measuring visual search performance evaluates the reaction time to the detection of a target amongst a set of distractor items [48, 199]. Many factors modulate search performance. Search targets require attributes that can guide attention to distinguish targets from distractors, such as colour, motion, orientation, and size (see Wolfe and Horowitz 2004 [201] for a complete list of attributes and their assessments). The efficiency of the search is also affected by the degree of similarity between the target and distractors [47]. Similarly, spatial layouts can impact search performance—targets that appear farther from center gaze [26] or crowding [188] will negatively affect results. Furthermore, searches for previously sought targets are faster [74]. Target memory has been found to extend to the previous seven trials [108]. However, when participants search repeatedly within a small set of objects, significant improvements in speed are not typically observed over the course of a session [202].

Within a coherently organized search space [200], people are expected to evaluate each item serially to determine whether it is the target. As such, reaction times can be predicted in serial, self-terminating search environments by set sizes [81, 119]. The Serial Search Model has been instrumental in predicting search time in time-sensitive and

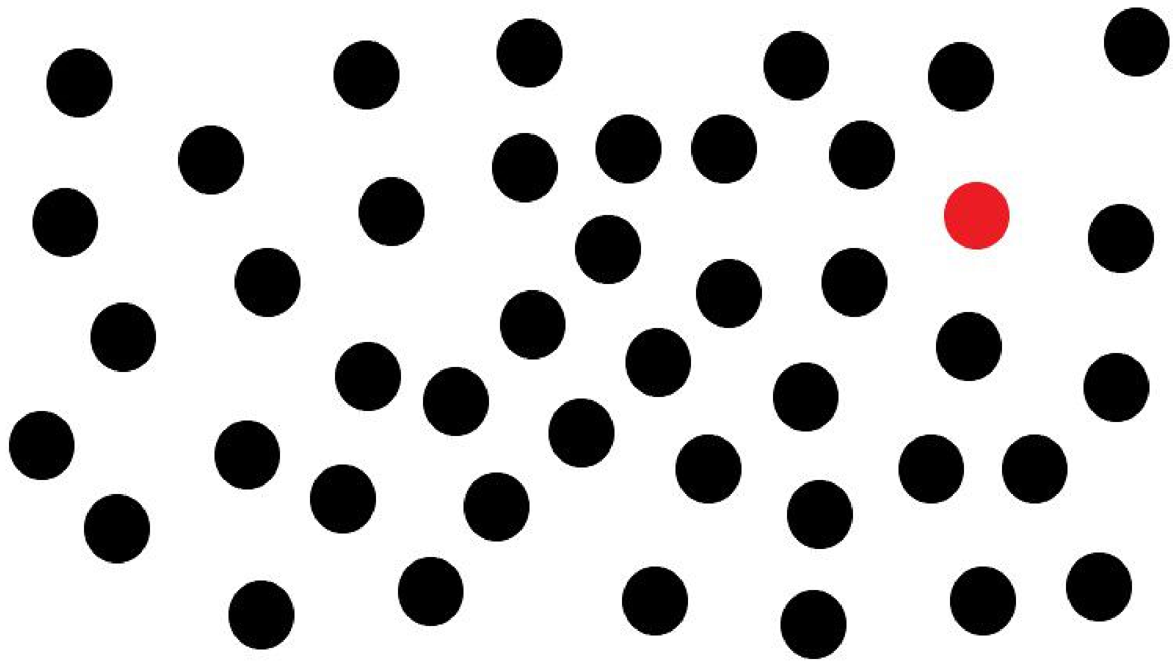


Figure 2.3: Red target amongst black distractor items.

linear environments [194, 57]. Nonetheless, the generalisability of the Serial Search model has its limits, especially when considering the effects of bottom-up and top-down attentional processes. For example, when target items "pop-out" (bottom-up attentional mechanisms), non-target items are bypassed and disregarded in the search process [205]. Furthermore, in the absence of visually distinctive targets, another limit exists in the form of expectancies (top-down attentional mechanisms). In a visual field of similar objects differentiated by colour, if we know to look for an orange object, we will only attend to the orange coloured ones [49, 130]. This forms the premise to Wolfe's Guided Search Model [198], which states that with pre-existing knowledge of the target's visual properties, attention can be directed to categorically similar items from the preattentive stage. It is because of this quality - that is, the expectancy of humans to use top-down attentional mechanisms when they are searching for items - that we will be designing our interfaces with a focus on the biological processes of perception rather than from a design perspective. We anticipate that as participants become more familiar with the study interfaces that they will be searching for target items using visual variables they remember the target items possessing such as specific colours or shapes.

2.2 The Design of Visual Icons

In the following sections, we review the design aspects of icons that informed our study interfaces' designs. First, we summarize common icon categorizations that will inform the classification of our study interfaces. Second, we review the visual characteristics of icons, and the effects certain properties impart. Finally, we report on recent design trends in icons and how they affect learnability.

2.2.1 Icon Meaning

Frameworks have long existed that taxonomize icons, often drawing inspiration from the field of semiotics (see Peirce [132]). The first serious discussion of categorizing icons in interfaces emerged in 1989 with Rogers' [141] proposal to identify icons based on form and function. In the years since, an increasing amount of literature has put forth additional terminology and classification. For example, to denote icons that lack any meaningful visual connection to the underlying concept they convey, terms have emerged such as arbitrary [141, 102, 105, 61] and sign [193]. Further classification can be observed in the approach to icons that represent their meaning conspicuously but contingent upon whether they reflect their underlying commands directly (representational [105, 19], concrete [137] etc.) or indirectly (abstract [137, 61], symbolic [141, 102, 193] etc.). Hybrid approaches that combine both direct and indirect representation of underlying function have been categorized as mixed [103] or associative [61].

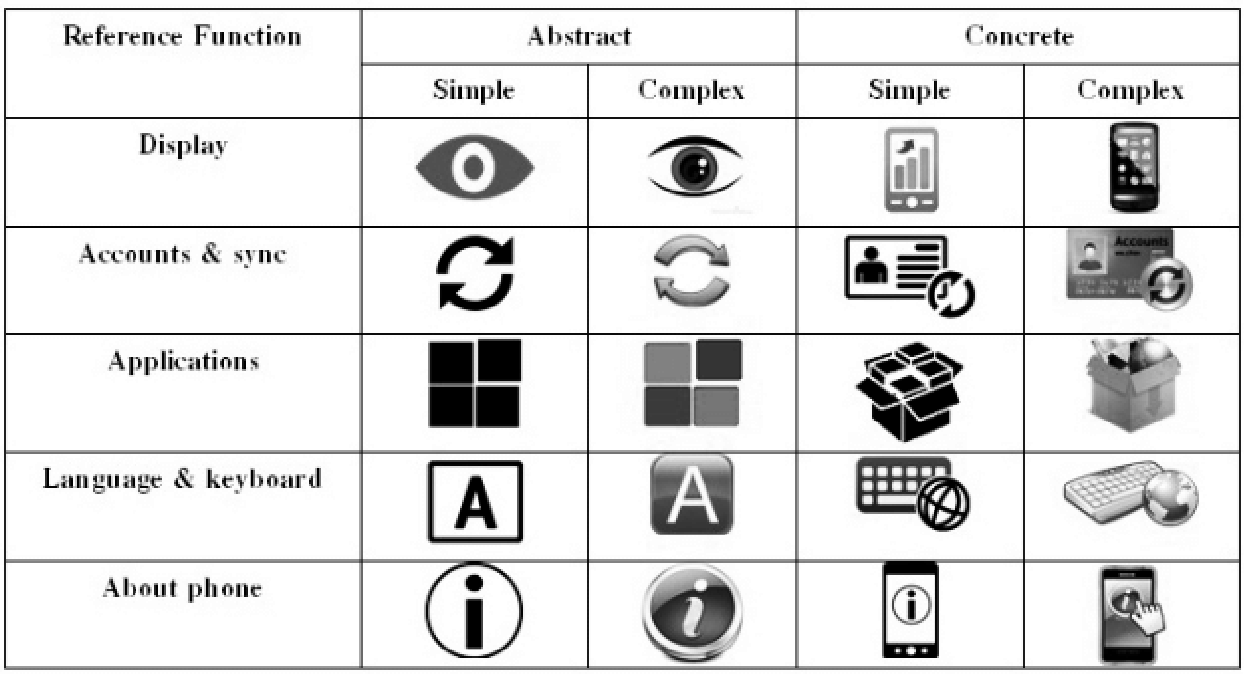


Figure 2.4: Four example icon sets of different styles and meaning. From "A study of the recognitions and preferences on abstract and concrete icon styles on smart phone from Easterners and Westerners' point of view" [29]. Copyright 2013 IEEE.

Among these categorizations, however, the success of making an icon distinguishable depends on how naturally it can depict its underlying function [19]. Research has suggested that ambiguity is not desired in icons [158, 145] and that icons that represent their underlying functions allow users to find icons faster [53]. Icons that do not naturally depict their underlying functions have been found to elicit a larger toll on neural processes to recognize and thus produce slower response times [78]. These icons are also discouraged as they suffer from major drawbacks such as context being culturally specific and thus would fare poorly across a wide demographic [53].

Furthermore, visual information has long been found to arouse emotions. Specifically, shapes and colour have shown to elicit emotional responses in the field of psychology and visual arts. Prior research into shapes indicates that

Figure 2.5: Example of flat interface. Visual variables such as colour, contrast,weight, and size are repeated across the entire set. Abstract (triangles dispersing outwards) and concrete (clock for timer) meaning can be observed in the interface. From "Icons Web Development Website Design Flat" [4].

humans greatly prefer curved shapes as it is associated with positive feelings [159, 192, 13]. Angular shapes, however, are associated with negative and threatening feelings. As such, common design guidelines advise the usage of circular shapes to convey comfort and approachability, triangular shapes to convey danger, and square shapes to convey security or robustness [176].

Similarly, colour has also been found to possess cognitive associations [95]. These associations are primarily driven by experiences and real-world artifacts that possess these colours (i.e. yellow stripes on bees and the colour yellow associated with warnings or danger) [171]. Colours such as red have shown to evoke feelings of excitement, green with nature, and blue with feelings of calm and competence [95]. Colours can have many associations depending on context and culture [75]. For example, yellow is sometimes associated with hatred in western hemisphere culture while in Chinese culture is associated with trustworthiness [52].

Comprehending icon categorization and how colour and shape evoke associations is an important step that informs the design of our study interfaces. While our primary objective is to observe the effects of visual variables on learnability in GUIs, icon interfaces cannot be evaluated independent of the effects of meaning. Understanding the differences in how meaningful an icon can be is important to eliminate unintended effects of meaning on results. In particular, two categorizations are of interest to us: concrete and abstract meaning. Abstract and concrete meaning can be observed in flat interfaces; a popular design language, flat design enforces a uniform design language where visual features such as colour, outlines, or material are repeated across an interface (see Figure 2.5 for contemporary examples). As such, while there are many ways icons have been categorized (e.g., arbitrary, sign, symbolic, representational etc.), we will however focus on concrete and abstract meaning due to their prevalence to enforce the ecological validity of our results.

2.2.2 Visual Distinctiveness in Graphical Icons

Visual distinctiveness—or the visual aspects that help differentiate graphical objects—in icons include colour, size, and shape [140]. Colour, one of the most prominent visual traits, can easily separate an icon from another. Despite humans' ability to see a considerable number of colours, most people can only differentiate and remember five to eight colours in a visual workspace [162]. One of the primary uses for colour in interactive systems is in highlighting items (e.g., searching [28, 31]). Size is another visible feature that makes icons distinguishable. Although a common use of the size feature is to create a cohesive interface (e.g., similar-sized icons used throughout a GUI; Figure 1.1 & 1.2), changing the size of an icon can make it distinct (e.g., MS Office [111] uses multiple sizes of icons).

Furthermore, the shape of an icon is a vital visual factor that represents the underlying meaning [123]. Shapes can make icons more easily discernible as people can identify far more shapes than colours [163]. However, shape distinctiveness, that is, the separability of shapes, is challenging to define precisely. Prior work has explored aspects of this concept: Julész [71] identified shape features that early visual systems detect, Burlinson et al. [22] proposed that open or closed shapes influence perceptual processing, and Smart et al. [161] investigated the perception of filled, unfilled, and open shapes in scatterplots. Researchers have also studied luminance [116] of icons for learnability. Studies have shown that abstract and ambiguous icons demand more cognitive processing to recognize [78] and often hinder users from quickly learning them. However, the primary question-—how visual variables of icons impact learning and recall—-remains unanswered.

Trends [212, 154, 65] in icon design have prioritized less-is-more (e.g. silhouette-style icons) and consistency (e.g. repetition of visual features such as colour) [169, 187]. This is notably demonstrated by flat design, with its minimalistic approach that reduces icons to essential visual features [154]. In flat design, extraneous visual elements such as three-dimensional effects, shadows, lights, and texture are removed. However, flat design is a highly contextual design language: it requires a knowledge of its affordances to successfully navigate, which can be a challenge for older generations [11, 100]. As visual embellishments are removed, inferential cues that allude to familiar real-world properties are eliminated, making flat design interfaces harder to understand and use [23, 40].

Unlike flat design, skeuomorphic—that is, "the use of imagery of real objects to represent a design object's function" [165]—design introduces real-world characteristics into digital representations that are not vital for functionality. It has been shown that skeuomorphic design is more intuitive [97, 100] and easy to use [128] than flat design because such designs are largely self-explanatory. While icons representing non-physical objects such as "Security Settings" won't necessarily benefit from a skeuomorphic approach over a flat approach, skeuomorphic design allows for more details which may benefit representations of physical objects. For example, disc jockeying applications retain a similar layout and icons akin to the physical turntables disc jockeys use due to the complexity of operating a turntable. The familiarity of the icons through skeuomorphic design helps make the interface more recognizable and usable to disc jockeys. However, skeuomorphic design suffers from clutter [11]; unnecessary affordances within the design may increase users' cognitive and visual load. Furthermore, the use of more "realistic" designs relies on a shared knowledge of the real-world metaphor [77]. For example, using a floppy disk as a symbol for saving files may be ineffective for younger generations that have never encountered a floppy disk in the real world. In empirical studies, preferences for skeuomorphic or flat design often fall along age lines. Studies have shown that digital natives [136] (individuals that came of age in a time of digital saturation) greatly prefer flat design [77, 203, 11] while digital immigrants [128] (those that have had to adapt to the digital world rather than having been born into it) prefer skeuomorphic design [30, 20, 209, 11].

Comparison studies of skeuomorphic and flat design often communicate unexpected results. For example, an eye-tracking study found that skeuomorphic design—an interface found to be easier to use [128]—had the longest average total task time, longest first to fixation, longest average fixation duration, and the highest average visit frequency amongst line-drawing, metro, flat, and skeuomorphic designs [204]. Xi et al. theorized that the complexity of the design impacts search efficiency. Interestingly, in a comparison between metro, line, flat, and skeuomorphic designs, Xi et al. found that abstract designs (metro and line) produced higher search efficiency, followed by flat design, whereas skeuomorphic design had the lowest search efficiency. Similarly, skeuomorphic design can lead to increased difficulty in executing web tasks and flat design can lead to increased speed of execution, according to research by Spiliotopoulos et al. [167]. Conversely, flat design has been found to increase cognitive load and introduce higher amounts of error than skeuomorphic design [23]. Furthermore, in a comparison between Windows 7 (skeuomorphic) and Windows 8 (flat) Schneidermeier et al. found that the former was more usable [154]. In summary, existing research does not support a definitive conclusion about the superiority of skeuomorphic or flat design.

Of the literature reviewed, Spiliotopoulos et al.'s findings are of particular interest to us. While we are not evaluating the differences between flat and skeuomorph interfaces, we will be investigating the learnability of visual features such as colour and shape in icons. However, as Spiliotopoulos et al.'s interfaces each had variations of colour, shadows, orientation, and shapes, it is difficult to establish whether the collective effect of all four visual features or particular variables contributed to their results. As such, we will proceed with evaluating the learnability of colours and shapes (visual features that have exhibited capacity to guide attention in visual search [201]).

2.3 Spatial Memory and Transition to Expertise

The ability to remember objects within a GUI is a hallmark of expertise which is imperative for a more efficient experience—command selection from recall is a faster and more fluid experience than visually searching an interface. Therefore, this chapter will briefly review spatial memory and how it affects performance in GUIs. For a more detailed treatment on spatial memory development, see Postma et al. [135] and Thorndyke et al. [175]. We also discuss the development of expertise in GUIs and the transition between modes of different performance ceilings.

2.3.1 Spatial Memory

The ability to retain the location of objects within one's surroundings (i.e., spatial memory) is developed through repeated interactions with those objects [50, 90, 104, 33]. Spatial memory has proven to be long-lasting [43] and large in capacity [85]. While typically developed simply as a byproduct of interactions with objects, it has been demonstrated that effort in learning item locations produces more effective spatial knowledge [35, 50].

Spatial memory is generally divided into two categories of spatial tasks: navigation and remembering object location in a static setting [107, 72]. The difference arises from the task's environment: while navigating, the viewpoint of the user is dynamic and only a portion of the space is visible; while in the task of remembering object locations, the space in a setting is static and entirely visible. Although spatial memory in navigation has some relevance in traversing GUIs [44], we will instead be focusing on spatial memory for object location.

Spatial memory has long been employed to improve user actions. Through repeated usage of an interface, users can quickly retrieve item locations from memory [33]. The ability to recall locations of commands from memory facilitates users' transition from novice to expert in GUI systems. This transition means the visual searches that are necessary as users familiarize themselves with interfaces become unnecessary once the spatial locations of objects within an interface are learned [152].

Spatial memory has been found to better facilitate revisitation efficiency in flattened command hierarchies compared to lists and menus. Researchers have tried to exploit spatial memory by laying out interfaces in spatially stable ways [34, 50, 146]. For example, Scarr et al.'s [147, 149] CommandMap showed a spatially stable icon arrangement in desktops, yielding better learning and recall of icons, even for real tasks [148], because users could leverage spatial memory [35, 147, 184]. Similarly, Gutwin et al. [67] and Cockburn et al. [32] showed that a stable layout displaying every available command could increase recall efficiency compared to hierarchical ribbons or menus. Spatially stable icons have been found to improve learning and recall in multi-touch tablets [60, 68, 69, 183], smartwatches [96], smartphones [208, 211], digital tabletops [186], and even in virtual reality [58].

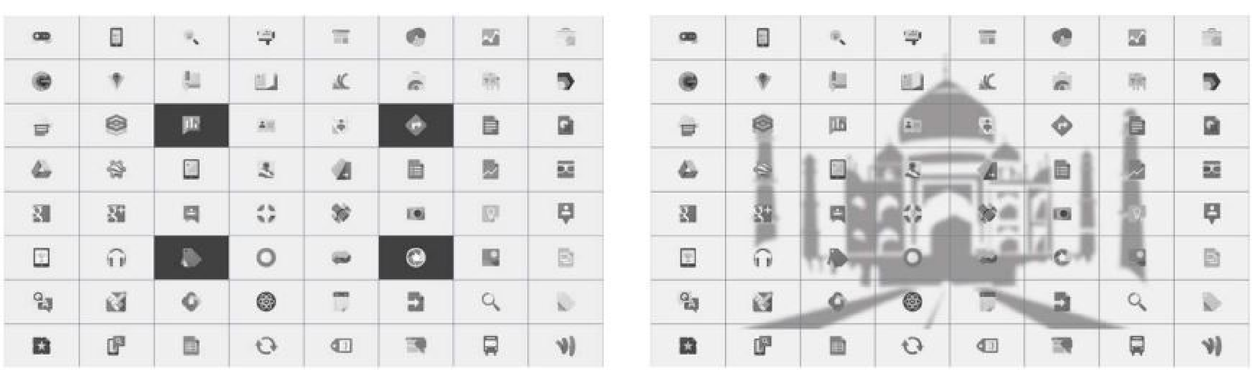


Figure 2.6: Example of "anchor points" (left) and an image of the Taj Mahal (right) as landmarks within an interface. Used with permission of ACM, from "The Effects of Artificial Landmarks on Learning and Performance in Spatial-Memory Interfaces" [184]; permission conveyed through Copyright Clearance Center, Inc.

Spatial memory benefits from landmarks [156, 157, 5]. For example, when looking for a book within a shelf, one starts in the topmost right corner, next to the book in red as they best recall. The right corner and the book in red serve as landmarks or points of references when searching for the book. They are identifiable features within a stable space that are distinguishable from their surroundings [106]. This concept extends to GUIs, wherein landmarks can be observed in defined areas such as screen corners or visual delimiters such as borders within an interface. Similar to their benefits in real life, landmarks have exhibited potential in GUIs to better facilitate memory-based command selections [6, 181, 182]. Researchers have exploited landmarks that are already present in the GUI environment. For example, the corners of a screen [69, 186, 155] can provide strong landmarks for icons near those locations. However, these natural landmarks often become useless in large interfaces (e.g., the middle area of a large screen) or a GUI with many icons because no landmark is present near those icons. In such cases, "artificial landmarks" [59, 184]—that is, landmarks made of "artificially created objects or colors" [182]—can aid learning and recall. Studies suggested that coloured blocks [6, 184], images as the background of a menu [184], and icons [185, 114] can be landmarks in GUIs to benefit spatial memory development.

2.3.2 Transition to Expertise

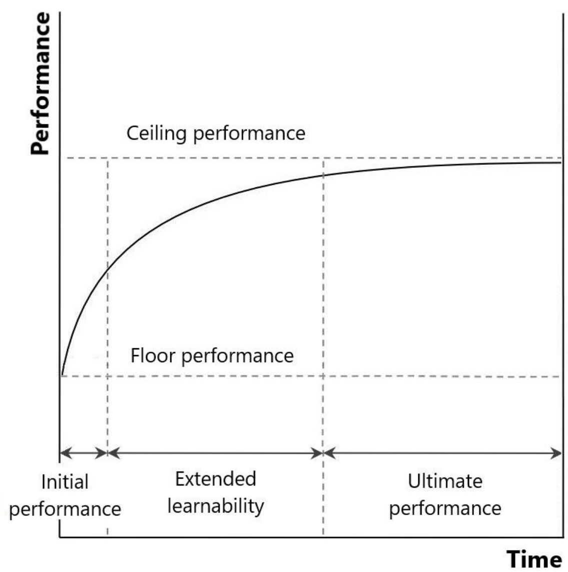


Figure 2.7: Initial performance represents the cognitive stage, extended learnability the associative stage, and ultimate performance the autonomous stage. Used with permission of ACM, from "Supporting Novice to Expert Transitions in User Interfaces" [34]; permission conveyed through Copyright Clearance Center, Inc.

The development of intramodal expertise—that is, the user's improvement over time within a single modality or interface mode such as WIMP UI—can be characterized by Newell and Rosenbloom's power law of practice [122]. Newell et al. suggests user performance improvement in a single interaction modality can be modeled using a power curve. This has been shown to apply to the context of human-computer interaction by Card's study, which observed user progression within a single modality over thousands of trials [24]. Per Anderson et al.'s [8] interpretation of the

Fitts and Posner stages of motor skill learning [55], the power law of practice curve can be subdivided into the three stages of skill acquisition: cognitive, associative, and autonomous (see Figure 2.7). In the case of GUIs, users learn the contents of an interface and visually look for commands in the cognitive stage. Second, in the associative stage, users already know the contents of the interface and begin to remember the commands in the UI. As a result, they can acquire those locations more quickly. However, users still perform local visual search after reaching the vicinity of a command in the associative stage. Last, in the autonomous stage, users can recall a command's location from memory and visit it without searching for it visually.

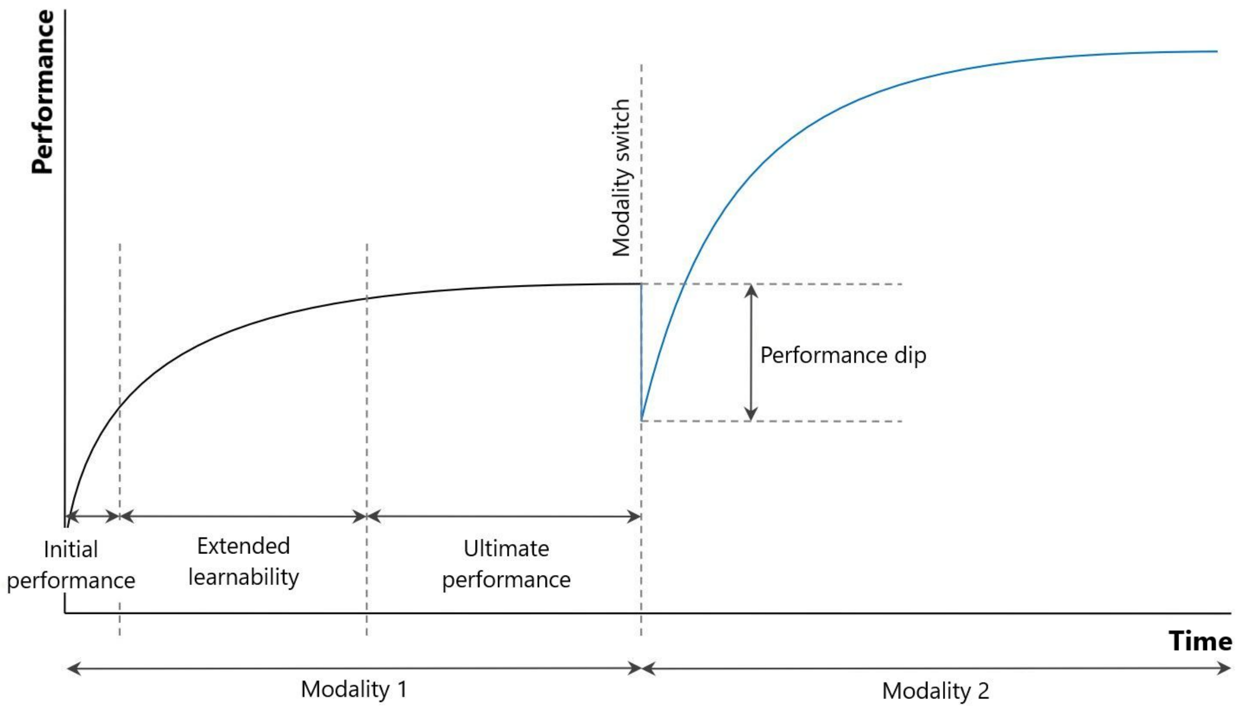


Figure 2.8: Performance curves of intra and intermodal user performance. Note the reduced performance at the initial transition to a different modality. Used with permission of ACM, from "Dips and Ceilings: Understanding and Supporting Transitions to Expertise in User Interfaces" [150]; permission conveyed through Copyright Clearance Center, Inc.

Intermodal expertise development happens when users transition between one modality to another typically with a higher performance ceiling [151]. Cockburn et al. characterizes the development of intermodal expertise as a combination of two power law curves [151]. An example of the development of intermodal expertise is the transition users make from WIMP UI to keyboard shortcuts. Expert interfaces in GUIs such as keyboard shortcuts possess higher performance ceilings (see Figure 2.8); these modalities typically provide greater amounts of executable commands versus traditional WIMP interfaces that are limited by available visual space [15]. However, expert interfaces are not as widely used as WIMP interfaces. The reason for the WIMP interface's popularity—that is, the ease of the 'see and point' quality of the interface—often traps users in this modality even once they have familiarized themselves with the interface [160]. Despite evidence of performance advantages that can be achieved by leveraging expert interfaces [125], users often do not move beyond point-and-click interactions [18, 27, 98] as they can be discouraged by the perceived loss of performance when making the switch [42]. Furthermore, a sense of “good enough” in regards to existing interactions discourage users from adopting expert interfaces or different strategies [160, 27]. Moreover, performance ceilings may present within a single modality. For example, the transition between recognition-based command selection—exhibited by visually searching an interface for the command—and recall based command selection where users are selecting commands within an interface from memory can also be represented by Scarr et al.’s overlaid power law curves (see Figure 2.8).

Due to the aforementioned majority of users that never move beyond point-and-clicking modalities, WIMP point and clicking) will be this study’s modality of choice. While we are conducting our study within a single modality, we will be observing the transition between recognition based and recall based command selection within a WIMP interface.

3 Characterising Visual Distinctiveness of Icons and Systems for General Study Methods

We carried out two studies to investigate the effects of visual distinctiveness of icons on learnability. We manipulated two visual variables—*shape* and *colour*. While shape and colour are not the only possible factors involved in icon distinctiveness, they were selected for their potential in visual search [201]—a perceptual task users undertake when seeking commands in a GUI. Although other factors such as motion, orientation, and size have been evaluated as leading guiding attributes in visual search [201], modern toolbars rarely employ motion outside of alerts and notices. With the prevalence of flat design, icons in toolbars often retain the same sizes and orientations. Furthermore, icon interfaces cannot be evaluated independent of the effects of meaning as the shape of an icon can also be representational; icons are designed to inherently represent tools or functions. As such, we also consider the cognitive variable of *meaning* in our studies as well [123, 19]. The following framework, methods, and system described will apply to all the studies presented within this thesis.

3.1 Framework of Factors

Meaning: The meaning of an icon refers to the concept or idea that the icon's image conveys. Icons in our studies varied by the types of underlying meaning they possess:

- *Meaningless:* Icon shapes have no connection to real-world objects or to their underlying commands (e.g., a grey square for the command "Settings").
- *Contextual:* Icon shapes are representational of underlying commands, but require interpretation if unfamiliar (e.g., a summation symbol for the command "Formula").
- *Familiar:* Icon shapes are pictorial and match their underlying command (e.g., an image of a calculator for the command "Calculator").

Shape Distinctiveness: In this work, we define levels of shape separability with respect to trends observed in modern icon design.

- *None:* Icons have no differences in shape (e.g. icons are all identical circles).
- *Medium:* Icons use different shapes, but are thematically uniform for visual consistency (similar sizes, weights, line styles, and borders).

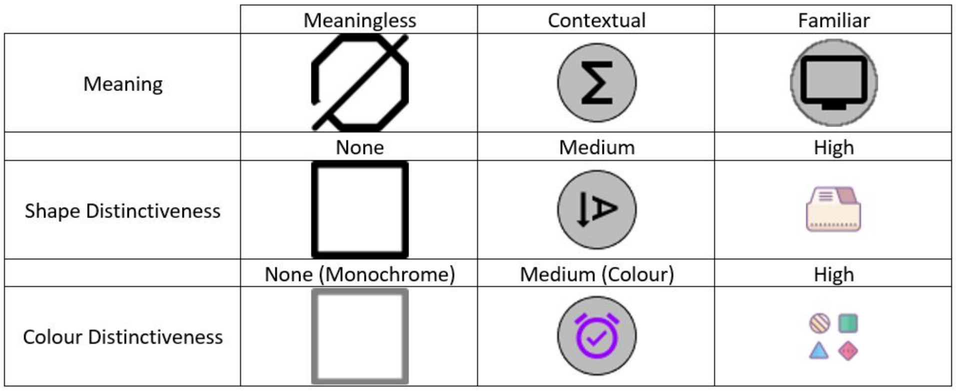


Figure 3.1: Examples of the framework levels.

Study	Condition	Meaning			Shape Distinctiveness			Colour Distinctiveness		
		Meaningless	Contextual	Familiar	None	Medium	High	None (Monochrome)	Medium (Colour)	High (Multi-colour)
S1	Concrete		x			x		x		
S1	Concrete+Colour		x			x			x	
S1	Mixed		x				x			x
S1	Abstract	x				x		x		
S1	Abstract+Colour	x				x			x	
S2	Square	x			x			x		
S2	Square+Colour	x			x				x	
S2	UnfamiliarShape	x					x	x		
S2	FamiliarShape			x		x		x		

Table 3.1: Icon properties of the interfaces in Study 1 & 2.

- *High:* Icon shapes are distinctly different from one another.

Colour Distinctiveness: We consider only basic levels of colour distinctiveness, because people's ability to distinguish colours is much lower than the ability to distinguish shapes [163].

- *None (Monochrome):* All icons use only a single colour.
- *Medium (Colour):* Different icons use different colours.
- *High (Multi-colour):* Icons use several different colours.

Both the visual and cognitive variables operate in the context of the spatial arrangement of the icons—the two studies and pre-study reported in the following sections investigate how different combinations of the levels and factors above (see Figure 3.1 for examples of each factors) affect users' ability to learn the spatial location of the icon corresponding to each command. For a summation of the levels of meaning, colour, and shape distinctiveness in each of the study interfaces, refer to Table 3.1.

3.2 Study Methods

The following description of study methods and tasks will be repeated across every study presented in this thesis. The experiment apparatus was conducted in the control room in the A-lab of the HCI research lab due to better lighting conditions (see Chapter 2.1.1 for the importance of well-lit viewing conditions). Participants were presented with ribbon-like menu interfaces that participants had to select prompted icons from over five blocks. Each interface consist of sixty icons in three rows to facilitate zero repetition of target trial locations between any of the interfaces. Target icon locations were repeated throughout each block. Targets were primarily located in the center of the screen to minimize the landmark effects of the edges of the screen as well as to position them to be better viewed by the center of the eye which perceives colour best (see chapter 2.1.1 for information regarding cone cell concentrations in the center of the eye). In the first study, five interfaces were evaluated; meaningless icons with (icons were one of many colours) and without (monochrome) colour, meaningful icons with and without colour, and an interface with multiple variations in colour as well as meaning. Condition orders are counterbalanced. Learnability was assessed through completion time, incorrect selections and hover amounts. Subjective responses and NASA TLX-style questionnaires were collected to gauge perceived performances and preferences.

3.2.1 Tasks and Stimuli

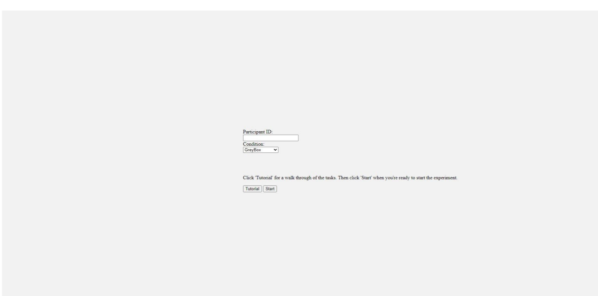


Figure 3.2: Participant ID input and condition selection screen.

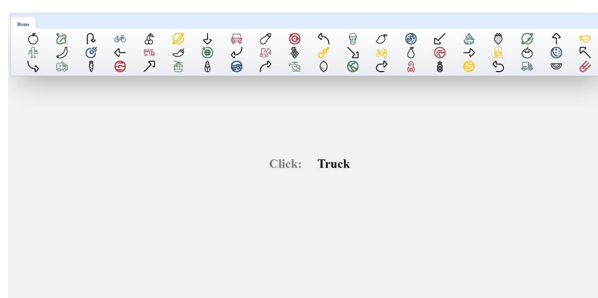


Figure 3.3: Tutorial interface.

The participant is first presented with a start page to input their participant id and condition (Figure 3.2) under the guidance of the experimenter. After inputting their participant id, participants first undergo a short tutorial to familiarize themselves with the system (Figure 3.3). The tutorial uses icons that will not appear in any of the study conditions. Following the completion of the tutorial, participants return to the start page where they will enter their id and select a condition to begin. Participants were instructed to complete the conditions as quickly and as accurately as possible. The study consisted of a series of trials in several interfaces, where each trial involved locating and selecting an icon. This task is commonly and frequently done in several toolbar-based or ribbon-based interfaces, such as Microsoft Word

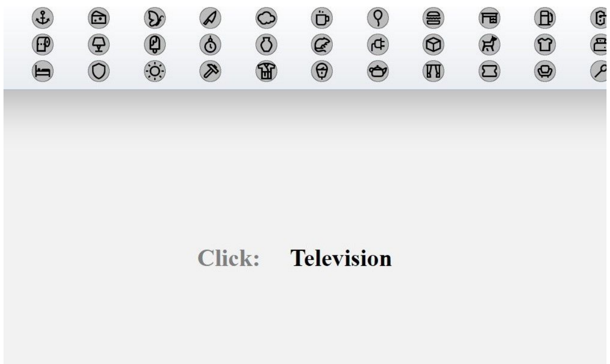


Figure 3.4: Example trial prompt. Participants are prompted to find the “television” icon. Icon locations remain static throughout the condition.

Figure 3.5: Hover tool-tip displaying the command name of an icon.

Figure 3.6: Icon flashing red indicating it was the incorrect selection.

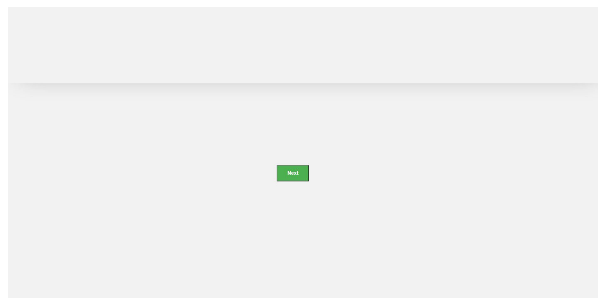


Figure 3.7: ”Next Trial” button prompt. The prompt appears at the correct selection of the target icon. The toolbar is covered while this prompt is up. Clicking ”Next” dismisses this button and starts the next trial.

Figure 3.8: ”Next Block” button prompt. The button appears at the end of the last trial of the last block in every condition. Pressing ”Next Block” prompts a SurveyMonkey link to appear for participants to answer questions about the condition.

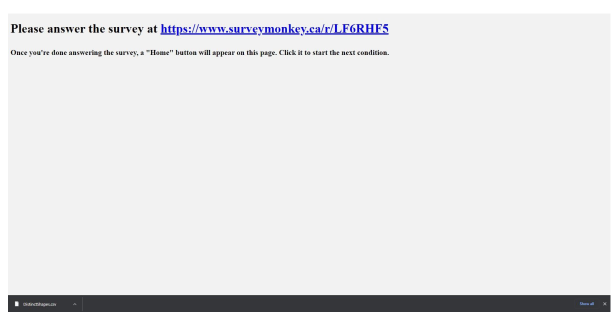


Figure 3.9: The SurveyMonkey link prompt that appears at the end of each condition.

2007 [111], Adobe Photoshop [3], or the GIMP graphics editor [173]. Every trial began by displaying a target word cue in the middle of a screen that remained visible for the entire trial, and participants were asked to find and select the corresponding icon from the toolbar (Figure 3.4). Participants could see the name of an icon as a tooltip after hovering over it for 300 ms (Figure 3.5). Each correct selection was indicated by a green flash at the selected location; red flashes were used to indicate incorrect selections (Figure 3.6). After selecting the correct icon, participants could proceed to the next trial by clicking on a "Next Trial" button that appeared in the middle of the screen (Figure 3.7). The button centred a participant's gaze and cursor position and started the timer of a trial. Between each block, participants had to press an orange "Next Block" button to proceed to the next block (Figure 3.8). The toolbar is obscured from view and the timer paused during the appearance of the green "Next" and orange "Next Block". For each interface, 9 out of the 60 icons were used as targets; these were sampled from three general areas of the toolbar [184]: 3 from the corner regions (first and last three columns), 4 from the edges (top and bottom rows) and 2 from the middle row. No target position was repeated among the five interfaces. Target positions in each interface were repeated across all participants in random order of appearance. After the completion of a block, participants are presented with a link to the SurveyMonkey survey of the condition they completed (Figure 3.9) and a button to return to the start page. Once the participant completes the survey, they will return to the start page and proceed to the next condition.

3.3 System for Experimental Evaluations

In the following section, we will review the system and apparatus used in our studies.

3.3.1 Apparatus

The study system (used in Studies 1, 2 and the pre-study) was written in JavaScript, HTML and CSS, and ran in the Chrome browser. Each condition is a separate html page, its icons laid out in vertical div grids of 3. Functionality is provided by a single javascript script that every html page shares. The study used a 27-inch monitor at 1920x1080 resolution, running on a Windows 10 PC with an Nvidia GTX 1080Ti graphics card. The system recorded all performance data; subjective responses were collected with SurveyMonkey.

3.3.2 System

A start function uses the condition and participant id from local storage to determine the correct array of target names. If arrays do not exist for hovers and clicks, arrays are initialized using headers (such as participant id, trial number, block number, click number etc.).

A shuffle function is called at the beginning of every block to shuffle the order of trial targets. The shuffle function uses Sattolo's implementation of the Fisher-Yates shuffle. If the first item in the array is the same as the last item in the previous block, the order of trial targets will be shuffled again.

A jQuery tooltip—enabled on hover—, displays names of icons when the mouse hovers on the icon. When users hover over an icon for longer than 300ms, a variable tracking the amount of hovers within a trial is updated and information containing participant id, condition, block number, trial number, target icon name, name of icon hovered over, and the hover duration is pushed to the hovers array.

When an image is clicked, the image id is retrieved and compared against the current trial target to verify if the selection was correct. Every click is recorded to an array along with information such as the participant id, condition, block, trial number, the correct target name, the duration of the trial, whether the selection was correct or incorrect, and the amount of clicks within the trial. If the selection was correct and was the last trial of the block but not the last block of the condition, a prompt is displayed to start the next block. If the selected icon was correct and was the last trial in the last block of the condition, the link for the relevant SurveyMonkey survey is displayed. If the selection was incorrect, the icon that was clicked will flash red for half a second.

4 Study 1

Effects of Colour in Meaningless and Contextual Icons

The goal of the first study was to evaluate the effects of colour and meaning, as the visually distinct features within an interface, on learnability in meaningless and contextual icons (see Chapter 3.1 for definitions to these levels). We provide a measure of the effects through repeated measurements of completion time, hover count, and incorrect clicks. The system described in 3.3 was used to conduct this study. A description of the interfaces used in the study follows as well as demographics and recruitment methods.

The efficacy of search within a set of distractors is influenced by visually separable differences between the target and distractor [47]. Furthermore, for features to effectively "guide" search, the feature differences between the target and distractor have to be greater than one noticeable difference [118]. As such, the icons in the interfaces in Study 1 possess at least two noticeable visual features such as colour or orientation—visual features reported as leading guiding attributes in visual search [201].

While there is no consensus on the benefits of non-spatial features in aiding the learnability of targets from distractors, we theorize that visually distinct features can be helpful when learning GUI layouts by allowing people to attend to objects that share the visual features of the target object—a phenomenon noted in Wolfe's Guided Search Model [198]—which, will reduce the time to learn their locations within an interface. Additionally, we also theorize that visual features can replicate landmarks within GUIs that aid people's navigation within an interface.

4.1 Study Methods

In the following section, we will review the interfaces used in Study 1, our study design, and provide a brief overview of the demographics of our participants.

4.1.1 Interfaces

We developed five custom web-based desktop icon selection interfaces, each consisting of sixty icons (44 px in size) arranged in three equal rows and presented in a standard ribbon-toolbar structure. All icon toolbars appeared at the top of the interface and allowed two types of mouse-based interaction: selection and hover. Names of the icons were not shown in the UI, but could be seen in a tooltip after hovering the mouse over an icon for 300ms. Icons were created using the GIMP image editor, using source images from freely-available icon sets such as material.io and icons8. We used five experimental interfaces in Study 1, described below and shown in Figure 4.1.

Concrete. The *Concrete* interface used monochrome icons similar to those found in standard mobile and desktop environments. The interface had contextual meaning, medium shape distinctiveness, and no colour distinctiveness (see Chapter 3.1 for definitions to these levels). The icons were chosen to avoid images of real-world objects, and therefore had contextual meaning. Although icons varied in shape, the level of distinctiveness was reduced by adding a circular grey background with a 1-pixel black border.

Concrete+Colour. The *Concrete+Colour* interface used icons similar to *Concrete* in terms of shape distinctiveness and meaning (no icons were repeated). The interface had contextual meaning, medium shape distinctiveness, and medium colour distinctiveness (see Chapter 3.1 for definitions to these levels). Icons were given a colour from a set of twelve unique colours; colours were equally distributed among the 60 icons. Colour brightness was adjusted to make icons with different colours clearly differentiable, and colours were not repeated for neighboring icons. The addition of colour provided the user with new landmarks that could be valuable for remembering locations (e.g., "it was the blue icon next to the red icon").

Abstract. The *Abstract* interface used meaningless monochrome icons consisting of circle and octagon shapes that were augmented with partial or full crossing lines, gaps in the outline, or dots in the centre of the icon. Icons in this set provided medium shape distinctiveness: each shape was different, but the set shared several basic visual properties. The interface had meaningless meaning, medium shape distinctiveness, and no colour distinctiveness (see Chapter 3.1 for definitions to these levels).

Abstract+Colour. The *Abstract+Colour* interface used icons that were similar in design to *Abstract*, but used a square base outline. Colours were added to icons as described above for the *Concrete+Colour* interface. The interface had meaningless meaning, medium shape distinctiveness, and medium colour distinctiveness (see Chapter 3.1 for definitions to these levels).

Mixed. The *Mixed* interface used icons with high shape distinctiveness (variations in size, shape, weight, and texture) and high colour distinctiveness (icons used a variety of colours). These variations provide users with two different types of landmarks to assist their location memory. The icons were adapted from a real-world set, and had contextual meaning. The interface had contextual meaning, high shape distinctiveness, and high colour distinctiveness (see Chapter 3.1 for definitions to these levels).

4.1.2 Procedure

At the beginning of the study session, participants completed an informed consent form and were given an overview of the study. After filling out a demographic questionnaire, participants completed a practice round consisting of 4 trials and 4 blocks with an icon set not used in the main study. They then completed 5 blocks of 9 trials for each of the five interfaces. The study followed a within-participant design, with the interfaces counterbalanced using a Latin square model. After each interface, participants completed NASA-TLX [70] questionnaires; after all interfaces, participants answered final questions about their preferences. Last, they reported their strategies for remembering target locations.

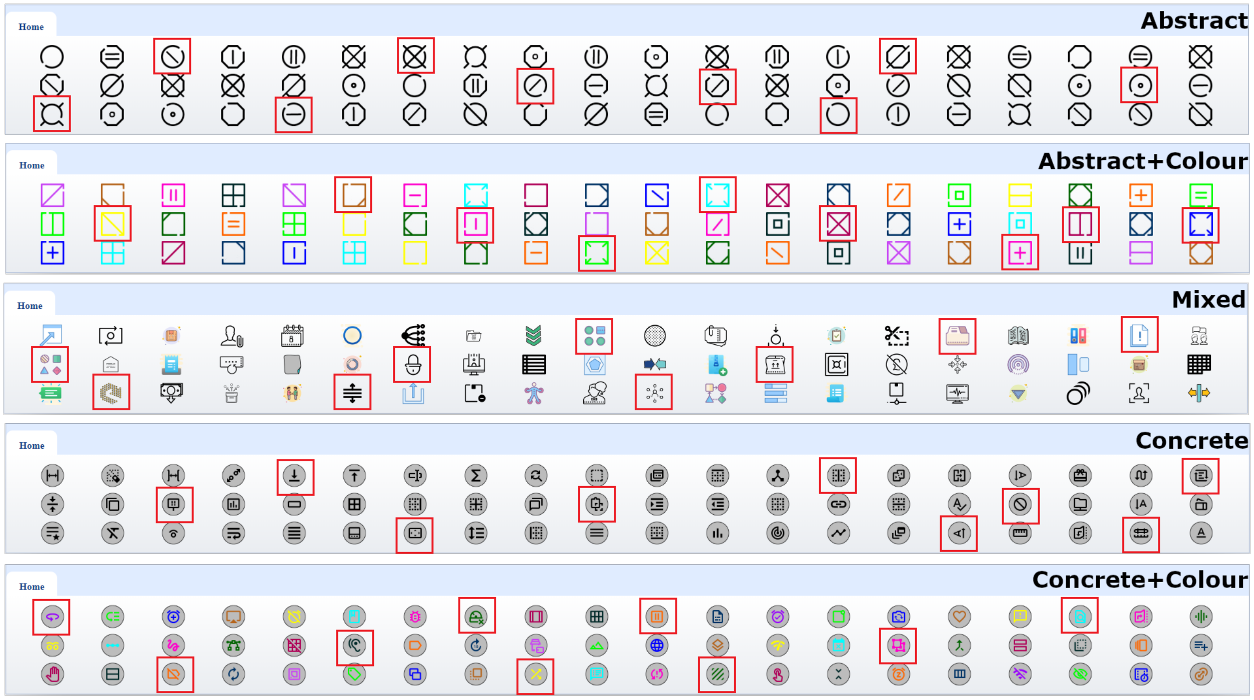


Figure 4.1: Screenshots of the five interfaces used in Study 1. Target objects are outlined in red.

4.1.3 Study Design

The following are our four hypotheses for Study 1:

- H1: Increased *colour distinctiveness* will reduce completion time and hover count *(Abstract and Concrete vs. Abstract+Colour and Concrete+Colour)*;
- H2: Increased *meaning* will reduce completion time and hover count *(Abstract and Abstract+Colour vs. Concrete and Concrete+Colour)*;
- H3: Increased *shape distinctiveness* will reduce completion time and hover count *(Mixed vs. Concrete+Colour)*;
- H4: Increasing both *colour distinctiveness* and *shape distinctiveness* will lead to a larger reduction in completion time and hover count *(Mixed vs. Concrete)*.

Our studies used a within-participants design in a series of planned comparisons using different sets of conditions. The dependent measures were:

- Completion Time: The duration in milliseconds from the appearance of a prompt to the selection of the correct icon;
- Hover Amounts: The duration in milliseconds spent hovering over an icon—an action that allowed the participant to see the name of the icon's command;
- Errors: The number of incorrect icon selections before the correct icon was selected.

For all studies, we report the effect size for significant RM-ANOVA results as general eta-squared: h^2 (considering .01 small, .06 medium, and $>$ .14 large [36]), and Holm correction was performed for post-hoc pairwise t-tests.

4.1.4 Participants and Recruitment

20 participants (10 men, 9 women, 1 non-binary), ages 20-44 (mean 26, SD 5.4), were recruited from a local university and received a $15 honorarium. All participants had normal or corrected-to-normal vision, and none reported a colour-vision deficiency. All participants were highly familiar with desktop and mobile applications (up to 10 hrs/wk [3 participants], 20 hrs/wk [4 participants], 30 hrs/wk [1 participants] and over 30 hrs/wk [12 participants]). The study took 90 minutes. 10 participants reported primarily issuing commands by navigating GUIs with mice and 10 reported using keyboard shortcuts. Overall participants were familiar with keyboard shortcuts (1-5 shortcuts [7 participants], 6-10 shortcuts [9 participants], 11-15 shortcuts [2 participants], 16-20 shortcuts [1 participant], and over 20 shortcuts [1 participant] regularly utilized).

4.2 Results

We report the effect size for significant repeated measures analysis of variance (RM-ANOVA) results as generalized eta-squared (h^2). Values $<$ .01 are considered small, $<$ .06 but greater than 0.01 as medium, and $>$ .14 as large [36]. Holm-Bonferroni corrections were performed for post-hoc pairwise t-tests.

4.2.1 Completion Time

Completion time was measured from the appearance of a word cue to the selection of a correct icon; no data was removed due to outlying values.

Mean completion times for the five icon sets are shown in Figure 4.2. Our first planned comparisons (H1 and H2) involved the effects of colour distinctiveness and meaning. A 2x2x5 RM-ANOVA (*Meaning X Colour Distinctiveness X Block*) showed effects of *Meaning* ($F_{1,19}$= 89.60, $p < 0.0001$, h^2= 0.54) and *Block* ($F_{1,19}$= 336.88, $p < 0.0001$, h^2= 0.71) on completion time, but no effect of *Colour Distinctiveness* ($F_{1,19}$= 2.99, p= 0.10). There were no interactions between the factors (all $p > 0.10$).

Follow-up tests for *Meaning* showed significant differences (all $p < 0.05$) between the concrete icon sets (Concrete and Concrete+Colour) and the abstract sets (Abstract and Abstract+Colour). Follow-up tests for *Block* showed differences between each successive pair except blocks 3 and 4.

Our third planned comparison (H3) used the Mixed and Concrete+Colour conditions to see whether shape distinctiveness would improve performance in icon sets that are already distinctive in terms of colour. However, a one-way ANOVA showed no difference ($F_{1,19}$= 0.086, p= 0.77).

Our fourth comparison (H4) used the Mixed and Concrete interfaces to see whether having two distinctive visual variables would improve performance (i.e., Mixed is more differentiable both in terms of colour and shape than Concrete). However, once again a one-way ANOVA showed no difference ($F_{1,19}$= 0.03, p= 0.86).

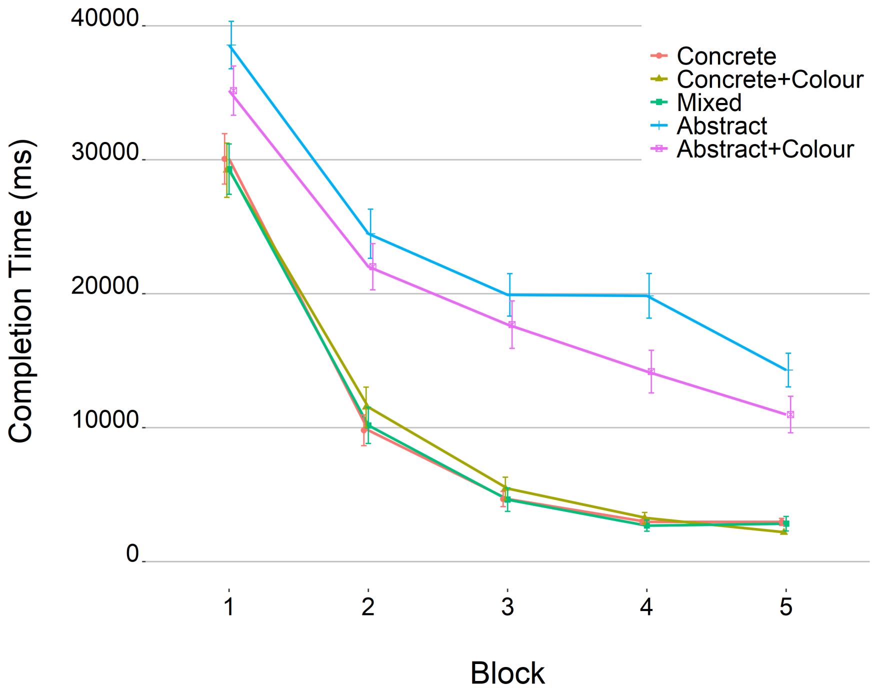


Figure 4.2: Mean completion time per block, by interface (s.e.).

4.2.2 Hovers

We measured hover count (where the participant held the mouse for 300ms over a target, showing the name) as a more sensitive measure of progress through the stages of cognitive, associative, and autonomous performance. As a participant moves from the cognitive to the associative stage, there should be a reduction in the number of icons that they need to inspect. Mean hover count per trial are shown in Figure 4.3. Results are very similar to those reported above for completion time: a 2x2x5 RM-ANOVA (*Meaning X Colour Distinctiveness X Block*) showed effects of *Meaning* ($F_{1,19}$= 117.5, $p < 0.0001$, h^2= 0.66) and *Block* ($F_{1,19}$= 353.65, $p < 0.0001$, h^2= 0.65) on hover count, but no effect of *Colour Distinctiveness* ($F_{1,19}$= 4.36, p= 0.051) (H1 and H2). There were also interactions between *Meaning* and *Colour* ($F_{1,19}$= 5.61, $p < 0.05$); as shown in Figure 4.3, the Abstract+Colour condition has fewer hover count than Abstract, whereas Concrete+Colour has more hover count than Concrete.

Follow-up tests for *Meaning* again showed significant differences (all $p < 0.05$) between both concrete icon sets (Concrete and Concrete+Colour) and both abstract sets (Abstract and Abstract+Colour). Follow-up tests for *Block* showed differences between successive pairs except for blocks 3 and 4.

4.2.3 Errors

We measured errors as the number of incorrect clicks before choosing the correct item. In some trials, participants clicked instead of hovering, leading to unusually high numbers of errors; we therefore removed 32 outliers out of 4500 total trials that were more than 3 s.d. from the mean. Overall errors were low (an average of 0.032 errors per trial). A

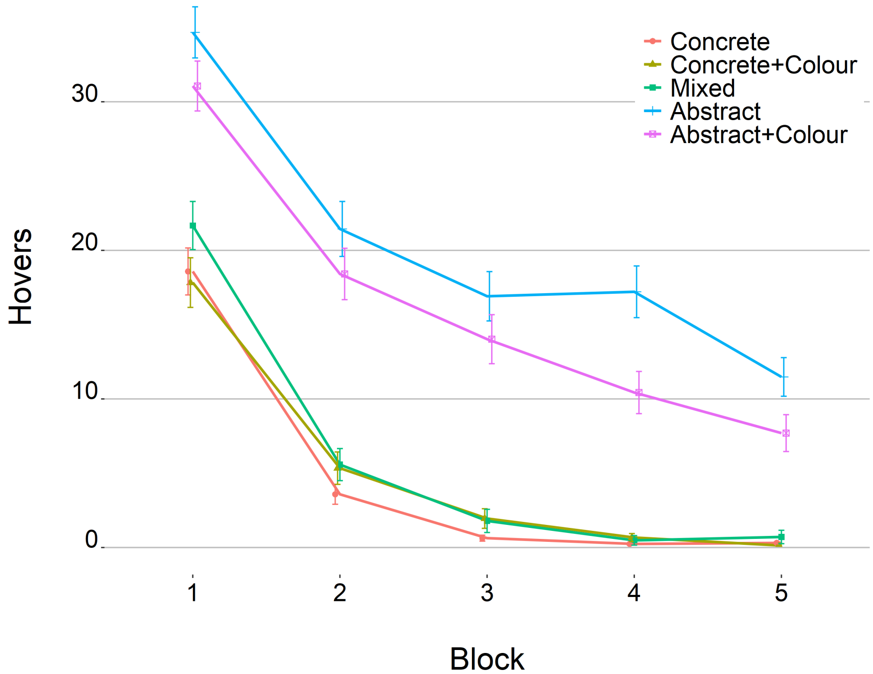


Figure 4.3: Mean hover count per block, by interface (s.e.).

2x2x5 RM-ANOVA (*Meaning X Colour Distinctiveness X Block*) to look for effects on errors showed a main effect of *Block* ($F_{1,19}$= 12.2, $p < 0.05$, h^2= 0.046) and a main effect of *Meaning* ($F_{1,19}$= 5.16, $p < 0.05$, h^2= 0.18). Follow-up t-tests showed that abstract icons had a significantly ($p < 0.05$) higher error rate (0.048 errors per trial) than concrete icons (0.018 errors per trial).

4.2.4 Subjective Responses and Comments

We used the Aligned Rank Transform [196] to perform RM-ANOVA on the NASA-TLX responses. As shown in Figure 4.4, mean scores of all TLX measures followed a trend similar to completion time. We found significant effects for all subjective measures. Follow-up t-tests revealed significant differences (all $p < 0.05$) between the two conditions with abstract icons (Abstract and Abstract+Colour) and the three conditions with concrete icons (Concrete, Concrete+Colour and Mixed) for every measure except physical effort. Significant effects were also found (all $p < 0.05$) in physical effort between Abstract and the three conditions with concrete icons as well as between Abstract and Abstract+Colour in perceived success.

Overall, participants preferred both Mixed and Concrete+Colour conditions. They also perceived them as the easiest and fastest conditions where they made the least errors. Pearson's Chi-squared tests revealed no significance (*Easiest*: X^2= 8.6, df= 4, p= 0.072; *Fastest*: X^2= 8.2, df= 4, p= 0.085; *Fewest Errors*: X^2= 8.2, df= 4, p= 0.085; *Preference*: X^2= 7, df= 4, p= 0.14). Results of the preference survey are summarized in Table 4.1.

Participants used a variety of techniques to learn and retrieve the icons. Eight participants stated that they relied on icon meaning and attempted to find a story or link to use as the basis for their memory: for example, one participant

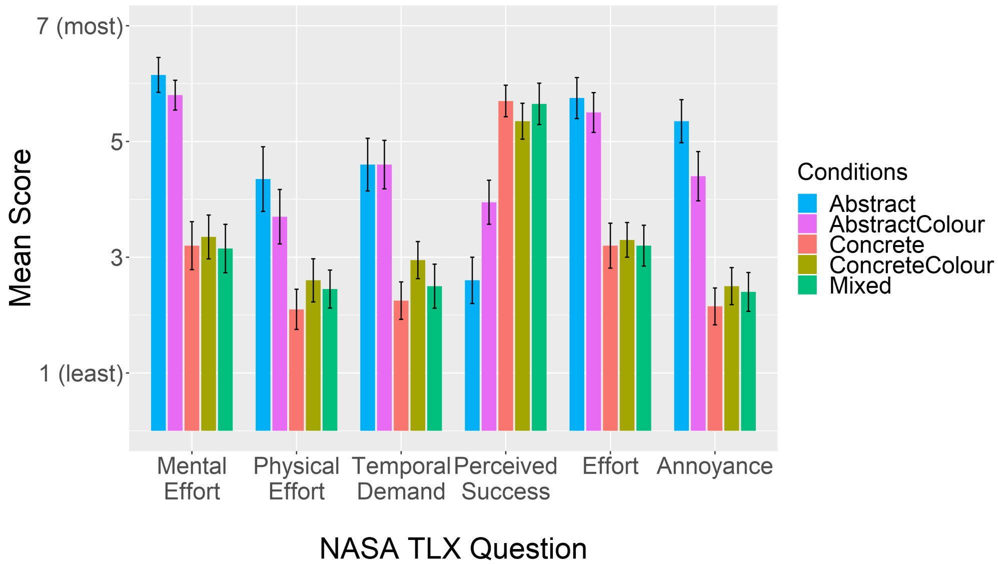


Figure 4.4: Mean NASA-TLX questions responses for Study 1 (s.e.).

	Easiest	Fastest	Fewest Errors	Preference
Abstract	0	0	0	0
Abstract+Colour	3	3	3	3
Concrete	3	3	3	4
Concrete+Colour	6	7	7	7
Mixed	8	7	7	6

Table 4.1: Summary of preference survey results.

said *"I tried to make a connection between the icon and the word."* Ten participants focused on remembering the spatial locations (at different levels of specificity); one stated *"[I recalled] the location of an icon if it was in the first, middle, or end [of the toolbar]."* Nine participants also commented on the value of shape distinctiveness. For example, a participant said *"If I had a good grasp of the icon's shape, it was easier to mentally place it in on the screen and find it again."* The same participant reported a challenge with the less-distinctive icon sets: *"I couldn't properly grasp a unique shape [in Abstract or Abstract+Colour], it became very difficult to mentally recall its position."* Finally, six participants also used the colour of icons; one stated *"colour added an additional element for memory."*

5 Colour Selection Study

In Study 1 exit questionnaires, when prompted to choose their preferred interface between colour and monochrome interfaces, participants overwhelmingly chose colour interfaces. However, contrary to expectations, results of colour as a factor on our dependent measures were not significant. One possible explanation for this result is that colour was presented in conjunction with additional visual variables that may supersede colour due to variances in participants' strategies for recalling icons. We also theorize that there may be too many colours to remember and it interfered with the usage of colour as a primary visual landmark for participants when recollecting icons.

In this small-scale study, we sought insight into the efficacy of colour as a visual landmark in ribbon interfaces by investigating whether colour, as the sole visual variable, improves learnability in comparison to a benchmark of no discernible visual variable. We also investigated the suitable number of colours to use in a condition by comparing learning rates with four, eight, and twelve colours.

5.1 Study Methods

This study was conducted using similar methods and procedures employed in Study 1 with a few modifications. In the following section, we will review changes made, the interfaces used, our study design, and provide a brief overview of the demographics of our participants.

5.1.1 Interfaces

We developed four custom web-based desktop icon selection interfaces for our study, each having 60 5-pixel border square icons. Due to the range that needed to be accommodated for twelve sufficiently distinct colours, we contend a few of the icons would fail the Web Content Accessibility Guideline's recommended minimum contrast ratio of 3:1. We chose four, eight, and twelve colours because of literature suggesting capacity limits in human short-term memory presenting within four [39] or between seven and nine [113] "chunks" of stimulus information. We additionally introduced a twelve colour condition to explore as per Alvarez and Cavanagh in their experiment [7]. Figure 5.1 provides examples of the four interfaces as well as the target icons in each interface. The four interfaces used in this study are described below:

Grey Boxes. The *Grey Boxes* interface repeated a single colour (grey) across sixty icons. *Grey Boxes* had meaningless meaning, no shape distinctiveness, and no colour distinctiveness (see Chapter 3.1 for the detailed framework).

Four Colours. The *Four Colours* interface distributed four colours (maroon, orange, green, and blue) evenly among sixty icons. No colours were repeated for neighbouring icons. *Four Colours* had meaningless meaning, no

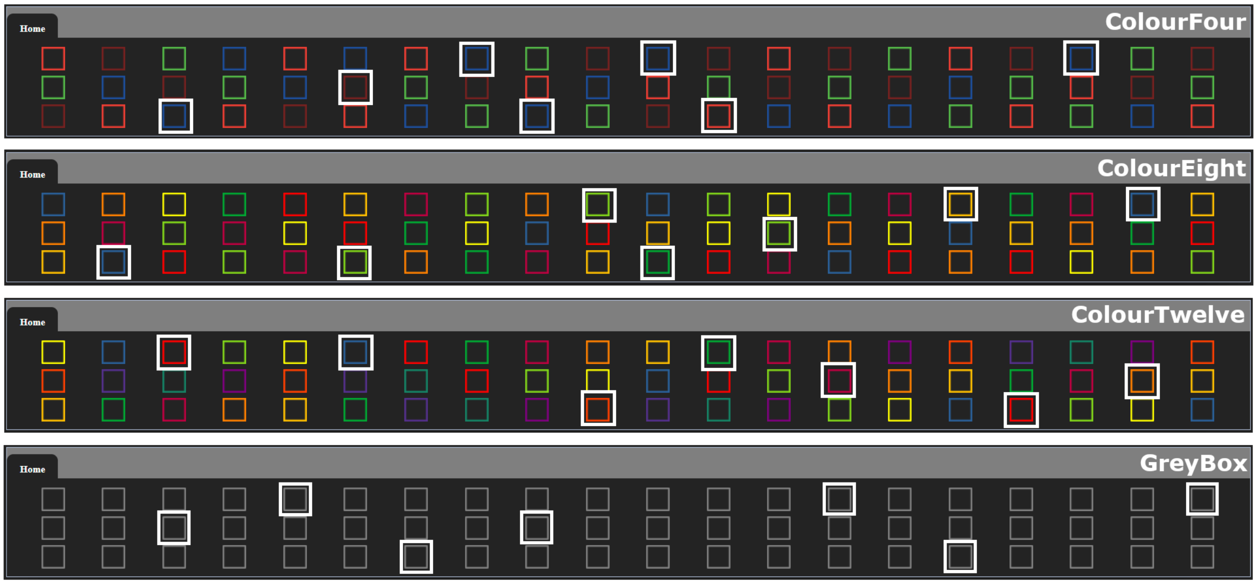


Figure 5.1: Screenshots of the four interfaces used in this study. Target objects are outlined in white.

shape distinctiveness, and medium colour distinctiveness.

Eight Colours. The *Eight Colours* interface distributed eight colours (red, yellow, green, blue, orange, gold, raspberry, and light green) evenly among sixty icons. No colours were repeated for neighbouring icons. *Eight Colours* had meaningless meaning, no shape distinctiveness, and medium colour distinctiveness.

Twelve Colours. The *Twelve Colours* interface distributed twelve colours (red, yellow, green, blue, orange, gold, raspberry, light green, dark green, dark orange, purple, and dark purple) evenly among sixty icons. No colours were repeated for neighbouring icons. *Twelve Colours* had meaningless meaning, no shape distinctiveness, and medium colour distinctiveness.

The twelve colours used in this study are represented in Figure 5.2.

5.1.2 Procedure

In the course of administering the first study, we observed engagement reducing towards the last thirty minutes of the study as fatigue set in. Participants reported meaningless icons as subjectively the most taxing. We took this into consideration as all the conditions in this study consisted of meaningless icons. First, trials per block were lowered from nine to seven. As demonstrated by previous graphs, inflection points in learning rates typically occur between the first and second block. The learning rate curves begin to taper in the fourth block and thus secondly, blocks were reduced from five to four. Third, in consideration of visual fatigue, we also explored using dark backgrounds to both increase visual comfort for participants [91] and to investigate whether visual fatigue had affected results in our previous study. It also benefited the visibility of the twelve separable colours more than a light coloured background—yellow on lighter backgrounds produced a contrast ratio lower than the darker colours on a darker background.

At the beginning of the study session, participants completed an informed consent form and were given an overview

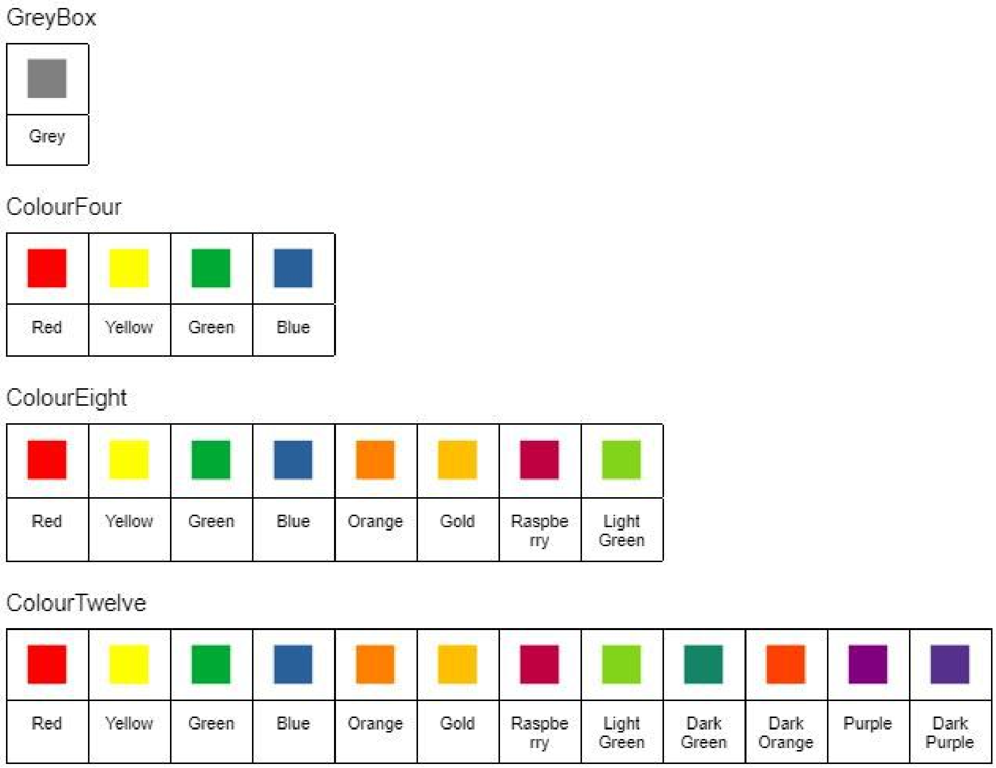


Figure 5.2: Samples of the colours used in each interface.

of the study. After filling out a demographic questionnaire, participants completed a practice round consisting of 4 trials and 4 blocks with an icon set not used in the main study. They then completed 4 blocks of 7 trials for each of the four interfaces. The study followed a within-participant design, with the interfaces counterbalanced using a Latin square model. After each interface, participants completed NASA-TLX [70] questionnaires; after all interfaces, participants answered final questions about their preferences. Lastly, they reported their strategies for remembering target locations. The study lasted approximately 60 minutes.

5.1.3 Study Design

The following are our two research questions for this study:

- RQ1: What *number of colours* will have the lowest completion time and hover count *(ColourFour vs. ColourEight vs. ColourTwelve*;
- RQ2: Increasing *colour distinctiveness* will reduce completion time and hover count *(GreyBox vs. ColourFour/ColourEight/ColourTwelve*.

Our studies used a within-participants design in a series of planned comparisons using different sets of conditions. The dependent measures were:

- Completion Time: The duration in milliseconds from the appearance of a prompt to the selection of the correct icon;

- Hover Amounts: The duration in milliseconds spent hovering over an icon—an action that allowed the participant to see the name of the icon's command;

- Errors: The number of incorrect icon selections before the correct icon was selected.

For all studies, we report the effect size for significant RM-ANOVA results as general eta-squared: h^2 (considering .01 small, .06 medium, and $>$ .14 large [36]), and Holm correction was performed for post-hoc pairwise t-tests.

Furthermore, we will also be introducing an additional factor for this small scale study called **number of colours**. In this work, we consider the number of colours present within an interface.

- *One:* Icons within an interface are all the same colour.

- *Four:* Icons are coloured one of four possible colours.

- *Eight:* Icons are coloured one of eight possible colours.

- *Twelve:* Icons are coloured one of twelve possible colours.

5.1.4 Participants and Recruitment

6 participants—a smaller number as this is intended to be a small scale study to give more clarity as to the number of colours to use—who did not take part in Study 1 (4 men, 2 women), ages 20-32 (mean 24, SD 4.8), were recruited from a local university and received a $10 honorarium. All participants had normal or corrected-to-normal vision, and none reported a colour-vision deficiency. All participants were highly familiar with desktop and mobile applications (up to 10 hrs/wk [1 participant], 30 hrs/wk [2 participants] and over 30 hrs/wk [3 participants]). 4 participants reported primarily issuing commands by using keyboard shortcuts and 2 reported navigating GUIs primarily with mice. Overall, participants were reasonably familiar with keyboard shortcuts (1-5 shortcuts [4 participants] and 6-10 shortcuts [2 participants] regularly utilized).

5.2 Results

We report the effect size for significant repeated measures analysis of variance (RM-ANOVA) results as generalized eta-squared (h^2). Values $<$ .01 are considered small, $<$ .06 but greater than 0.01 as medium, and $>$ .14 as large [36]. Holm-Bonferroni corrections were performed for post-hoc pairwise t-tests.

5.2.1 Completion Time

Figure 5.3 presents the median and range of completion time for each condition. No data was removed due to outlying values.

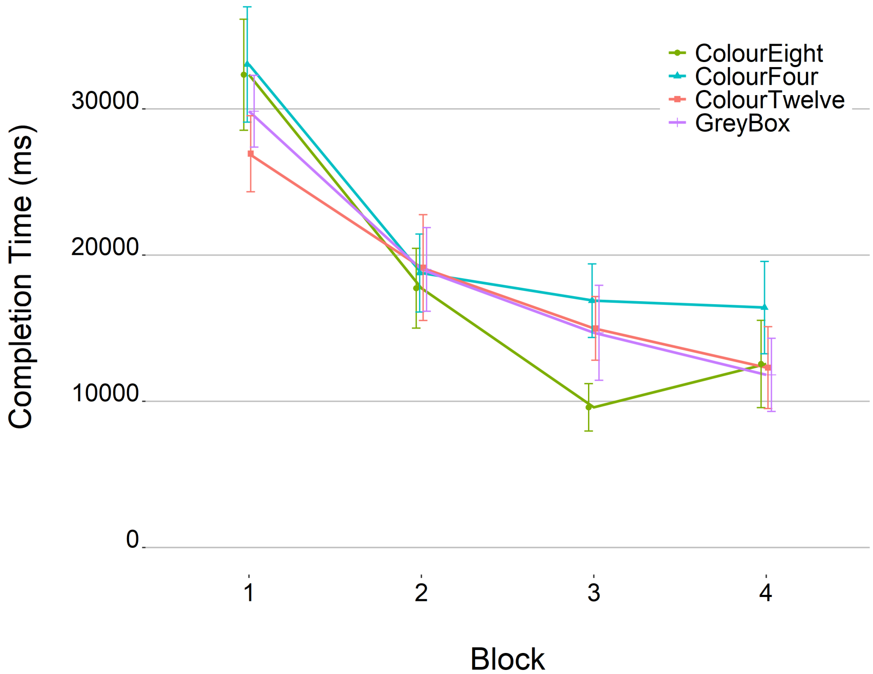


Figure 5.3: Mean completion time per block, by interface (s.e.).

In our first planned comparison (RQ1), a 3x4 RM-ANOVA (*Number of Colours X Block*) showed effects of *Block* ($F_{1,5}$= 115.96, p < 0.001, h^2= 0.6) on completion time but no effects of *number of colours* ($F_{2,10}$= 0.59, p= 0.57). Follow-up tests for *Block* showed differences between the first and every other block (all p < 0.0001).

In our second planned comparison (RQ2), a 4x4 RM-ANOVA (*Increasing Colour Distinctiveness X Block*) showed effects of *Block* ($F_{1,5}$= 63.58, p < 0.001, h^2= 0.6) on completion time but no effects of increasing *colour distinctiveness* ($F_{3,15}$= 0.4, p= 0.75). Follow-up tests for *Block* showed differences between the first and every other block (all p < 0.0001) as well as between the second and every other block (p < 0.05).

5.2.2 Hovers

In Figure 5.4, mean hover count per trial again reduces as participants transition into the associative stage. There was no evidence that the *number of colours* (RQ1) had an influence on hover count ($F_{2,10}$= 0.55, p= 0.6) in a 3x4 RM-ANOVA (*Number of Colours X Block*). However, similar to completion time, *Block* showed effects on hover count ($F_{1,5}$= 62.13, p < 0.001, h^2= 0.58). Follow-up tests for *Block* showed differences between the first and every other block (all p < 0.0001).

In our second planned comparison (*Increasing Colour Distinctiveness X Block*), a 4x4 RM-ANOVA showed effects of *Block* ($F_{1,5}$= 42.1, p < 0.001, h^2= 0.57). However, increasing *colour distinctiveness* did not affect hover count ($F_{3,15}$= 0.38, p= 0.77). Follow-up tests on *Block* showed differences between the first and every other block (all p < 0.0001) as well as between the second and every other block (all p < 0.0.5).

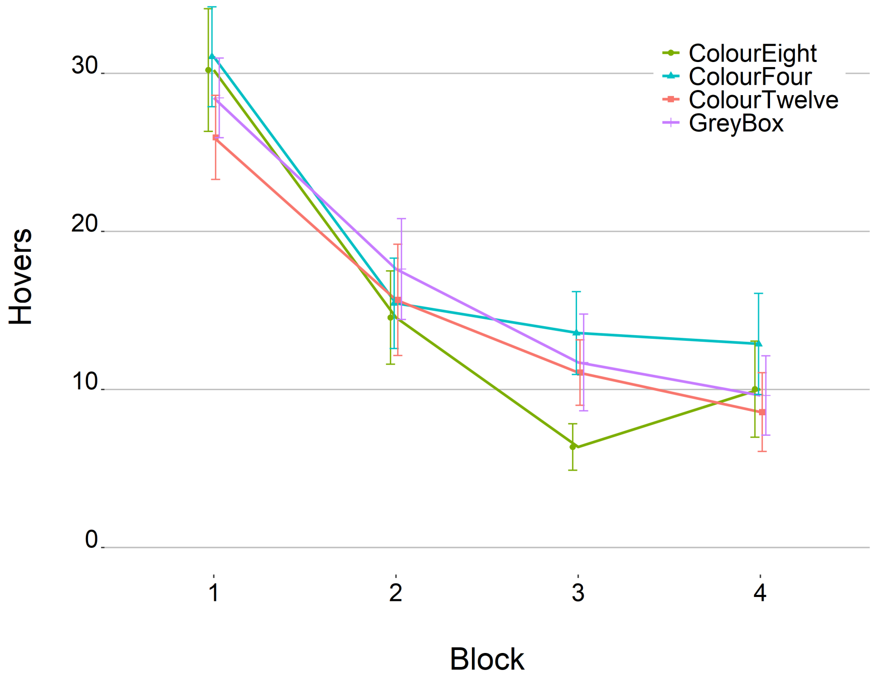


Figure 5.4: Mean hover count per block, by interface (s.e.).

5.2.3 Errors

Overall errors were low with an average of 3.87 errors per trial. RM-ANOVA showed no main effect of any of our main factors in our planned analyses (*RQ1 Number of Colours x Block:* $F_{2,10}$= 0.37, *p*= 0.7; *Block:* $F_{1,5}$= 2.4, *p*= 0.18; *RQ2 Increasing Colour Distinctiveness x Block:* $F_{3,15}$= 1.02, *p*= 0.41; *Block:* $F_{1,5}$= 1.47, *p*= 0.28).

Interestingly, despite possessing no unique visual indicators—every icon was a grey box—, GreyBox had the lowest mean errors amongst the four interfaces (1.86 errors per trial).

5.2.4 Subjective Responses and Comments

The following chart in Figure 5.5 summarises the mean scores of the collected subjective responses. Overall, participants scored favorably towards ColourFour. Looking at Figure 5.5, it is apparent that the results closely resemble all other dependent measures. We conducted an RM-ANOVA analysis using the Aligned Rank Transform [196] and found no significant effects for all subjective measures.

Respondents were asked to reflect on their methodology for learning icons. Four participants reported using colour to aid their learning of icon locations, predominantly by associating colours with stories or familiar items. One participant mentioned *"I associated the colours with situations, similar to previous parts. For example, clouds with sleep mode"*. Another participant improved upon this methodology by incorporating location into their memory strategies: *"Sleep mode: Blue is a sleepy colour AND sleep is in the middle of my life, so was easy to remember as a blue one in the middle"*. Two participants reported prioritising remembering location before colour (*"First, trying to get the*

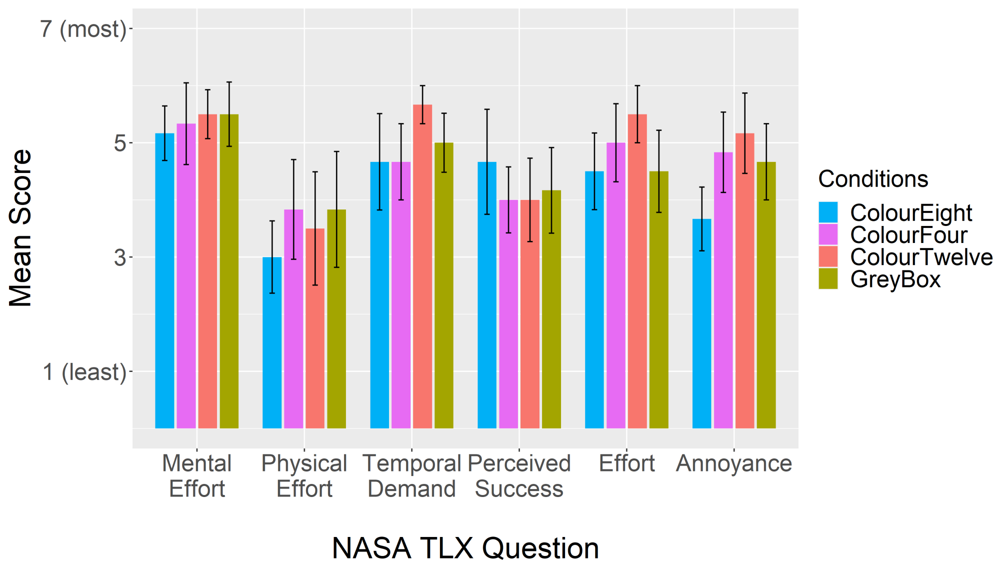


Figure 5.5: Mean NASA-TLX questions responses for the study (s.e.).

general area memorized. Then, associating something with the colour, such as red and getting a password wrong").

5.3 Discussion

We intended for this study to determine whether varying levels of medium colour distinctiveness could affect learnability of icons. However, in the course of this and the previous study, we were presented with contrary results suggesting colour, and shapes, in isolation produces no discernible difference on learnability in comparison to identical grey squares. To determine the validity and generalizability of our results—that is, contrary to our assumptions, colour does not improve learnability in graphical user interfaces— a second, larger study was planned with two of the interfaces from this study.

In informal conversations with participants at the end of their sessions, half expressed preference for four over eight and twelve colour icon sets. Participants had conveyed generally finding ColourFour easier to remember icons with and fastest to complete. Due to the lack of statistically significant difference between four, eight, and twelve colours, we moved forward with four colours. As more colours are introduced, the likelihood of having some colours look similar to another also increases thereby causing potential increased mental load within participants to discern between icons. With colour as the sole visual variable, fewer possible variations in colour would allow for perceptibly better separated icons which is an important determinant in attention per Itti et al. [84] and recall [116]. The background colour will be reverted to white to account for better contrast against a planned interface of monochrome icons.

6 Study 2

Effects of Colour, Meaning, and Shape as Sole Visual Variables

Study 1 suggested that colour did not improve learnability, and that icons with contextual meaning were substantially easier to learn than meaningless icons. A small-scale study further suggested that colour does not improve learnability (see chapter 5). However, a small-scale study with 6 participants lacks statistical power thus we conducted another study with 20 participants. In Study 2, we expand on these results and go into more detail on two questions: first, we verify whether colour does or does not improve learnability when it is the *only* visual variable (i.e., the icons have no shape distinctiveness); and second, whether it is the distinctiveness of an icon's shape or the meaningfulness of the image that assists learning.

6.1 Study Methods

Study 2 followed a similar method to study 1, but with two alterations. To reduce the overall time needed for the session, we reduced the number of targets from nine to seven, and the number of blocks from five to four (Study 1 showed clear learning effects within four trial blocks, see Figure 4.2). All other elements of the study method, procedure, and apparatus were identical to Study 1. In the following section, we will review the interfaces used in Study 2, our study design, and provide an overview of demographics of our participants.

6.1.1 Interfaces

The interfaces in Study 2 used a similar spatial layout of 60 icons as in Study 1, but used four new icon sets to explore our new questions about the effects of colour, shape distinctiveness, and familiarity (see Chapter 3.1 for the detailed framework). Figure 6.1 provides examples of the four interfaces as well as the target icons in each interface. The four interfaces used in this study are described below:

Square. The *Square* interface's icons were identical squares with a grey 5-pixel border. These icons have no colour differentiability, no shape differentiability, and no meaning. Therefore, the only way that participants could remember the correct icon was by remembering its spatial location. *Square* had meaningless meaning, no shape distinctiveness, and no colour distinctiveness.

Square+Colour. The *Square+Colour* interface used the same square shapes as *Square* for all icons, but the icons

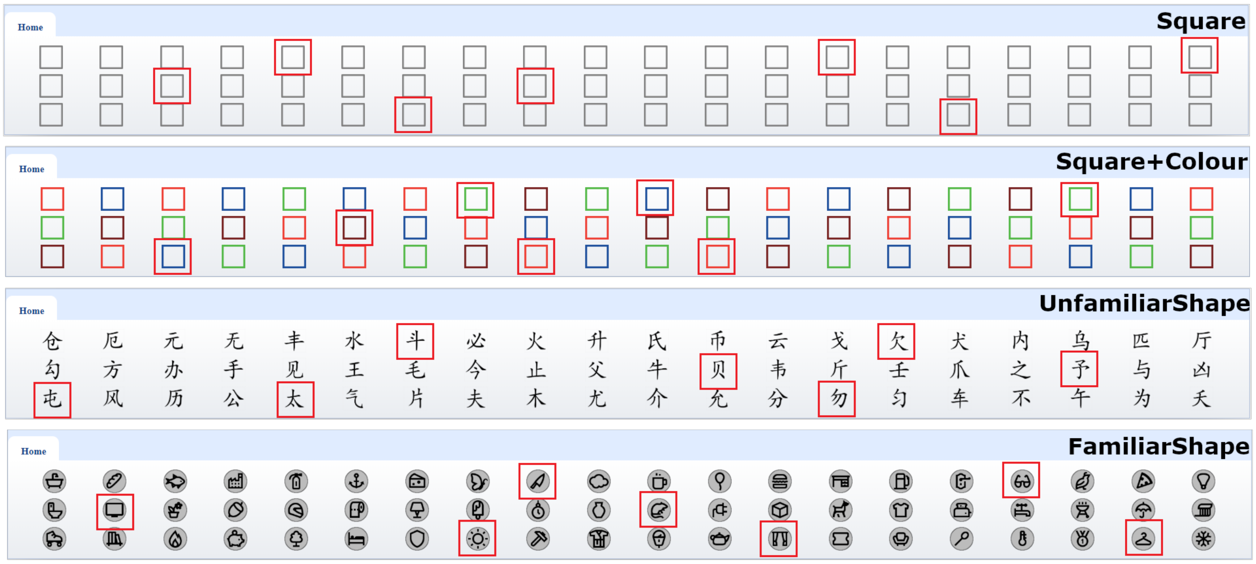


Figure 6.1: The four icon sets used in Study 2. Targets are outlined in red.

were coloured with one of red, green, brown, or blue. Colours were evenly distributed across the 60 icons, and no neighboring icons repeated a colour. Colour brightness was adjusted to maximize differentiability following Arthur et al. [10]. With no shape distinctiveness in the icon set, the colours provide additional landmarks for users to remember locations. *Square+Colour* had meaningless meaning, no shape distinctiveness, and medium colour distinctiveness.

UnfamiliarShape. The *UnfamiliarShape* interface showed monochrome four-stroke Chinese characters as icons. These icons had high shape distinctiveness (all icons were clearly different shapes). Chinese characters are meaningful, but only if the user is familiar with them—and our participants were chosen such that none knew these characters. Therefore, this icon set had no meaning for our study. *UnfamiliarShape* had meaningless meaning, high shape distinctiveness, and no colour distinctiveness.

FamiliarShape. The *FamiliarShape* interface used meaningful icons with imagery of recognizable real-world objects (Figure 6.1). Shape distinctiveness was medium, because we equalized several other visual variables such as size, line weight, and background shape (a grey circle with a 1-pixel black border). *FamiliarShape* had familiar meaning, medium shape distinctiveness, and no colour distinctiveness.

Icons were created using GIMP. FamiliarShape's images were sourced from material.io [2] and icons8 [1].

6.1.2 Procedure

At the beginning of the study session, participants completed an informed consent form and were given an overview of the study. After filling out a demographic questionnaire, participants completed a practice round consisting of 4 trials and 4 blocks with an icon set not used in the main study. They then completed 4 blocks of 7 trials for each of the four interfaces. The study followed a within-participant design, with the interfaces counterbalanced using a Latin square model. After each interface, participants completed NASA-TLX [70] questionnaires; after all interfaces, participants answered final questions about their preferences. Lastly, they reported their strategies for remembering target locations.

The study lasted approximately 60 minutes.

6.1.3 Study Design

The following are our four hypotheses for Study 2:

- H1: Increasing *shape distinctiveness* from zero (i.e. *Square* and *Square+Colour*) to high (i.e. *FamiliarShape* and *UnfamiliarShape*) will reduce completion time and hover count *(Square and Square+Colour vs. FamiliarShape and UnfamiliarShape)*;

- H2: Increasing *colour distinctiveness* in icons with no shape distinctiveness will reduce completion time and hover count *(Square vs. Square+Colour)*;

- H3: Increasing *familiarity* will reduce completion time and hover count *(UnfamiliarShape vs. FamiliarShape)*;

- H4: Even in icons without meaning, increasing *shape distinctiveness* will reduce completion time and hover count *(Square vs. UnfamiliarShape)*.

Our studies used a within-participants design in a series of planned comparisons using different sets of conditions. The dependent measures were:

- Completion Time: The duration in milliseconds from the appearance of a prompt to the selection of the correct icon;

- Hover Amounts: The duration in milliseconds spent hovering over an icon—an action that allowed the participant to see the name of the icon's command;

- Errors: The number of incorrect icon selections before the correct icon was selected.

For all studies, we report the effect size for significant RM-ANOVA results as general eta-squared: h^2 (considering .01 small, .06 medium, and $>$.14 large [36]), and Holm correction was performed for post-hoc pairwise t-tests.

6.1.4 Participants and Recruitment

Twenty participants who did not take part in Study 1 or the pre-study (16 women, 3 men, and 1 non-binary; ages 18-37, mean 24, SD 5) were recruited from a local university to complete the 60-minute study. They each received a $10 honorarium. Participants had normal or corrected-to-normal vision with no reported colour-vision deficiencies, and all were highly familiar with desktop and mobile applications (up to 10 hrs/wk [3 participants], 20 hrs/wk [3 participants], 30 hrs/wk [6 participants] and over 30 hrs/wk [8 participants]). 7 participants reported primarily issuing commands by navigating GUIs with mice, 11 reported using keyboard shortcuts, 1 reported using both and one reported using a trackpad. Overall participants were familiar with keyboard shortcuts (1-5 shortcuts [9 participants], 6-10 shortcuts [6 participants], 11-15 shortcuts [3 participants], 16-20 shortcuts [1 participant], and over 20 shortcuts [1 participant] regularly utilized). None of the participants could read Chinese characters.

6.2 Results

We report the effect size for significant repeated measures analysis of variance (RM-ANOVA) results as generalized eta-squared (h^2). Values $< .01$ are considered small, $< .06$ but greater than 0.01 as medium, and $> .14$ as large [36]. Holm-Bonferroni corrections were performed for post-hoc pairwise t-tests.

6.2.1 Completion Time

Mean trial completion times are summarized in Figure 6.2. No data was removed due to outlying values. We carried out analyses for each of our four planned comparisons.

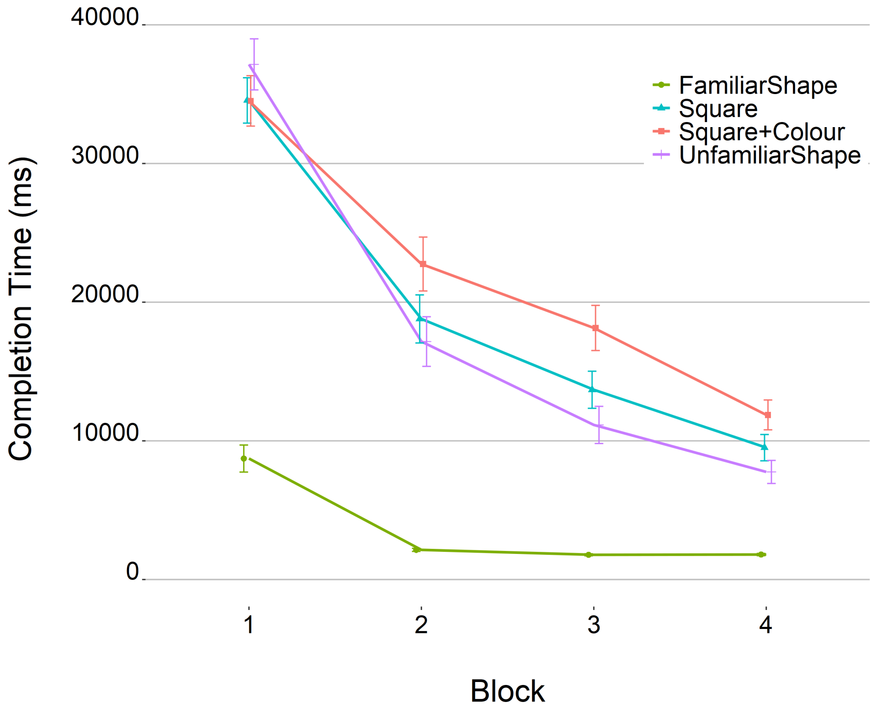


Figure 6.2: Mean completion time per block, by interface (s.e.).

First (H1), a 2x4 RM-ANOVA (*Shape Distinctiveness X Block*) showed effects of both *Shape Distinctiveness* ($F_{1,19}$= 124.22, $p < 0.0001$, h^2= 0.67) and *Block* ($F_{1,19}$= 181.67, $p < 0.0001$, h^2= 0.84) on completion time, as well as an interaction between the two factors ($F_{1,19}$= 12.44, $p < 0.01$, h^2= 0.09).

The effect of *Shape Distinctiveness*, however, must be considered in light of our third planned comparison (H3) of the familiarity of icon imagery—that is, in light of the large performance difference between the two interfaces with distinctive shapes. These interfaces (UnfamiliarShape and FamiliarShape) differ in terms of the familiarity of the icon imagery, and a one-way RM-ANOVA showed a highly significant difference between them ($F_{1,19}$= 112.24, $p < 0.0001$, h^2= 0.70). As can be seen in Figure 6.2, the UnfamiliarShape interface was much closer in learning rate to the two interfaces with square icons, and t-tests showed no significant differences between UnfamiliarShape and Square (p > 0.1), but showed that FamiliarShape was significantly different from all three other interfaces (all p < 0.001). In our results, therefore, the benefit of shape distinctiveness arose only when those shapes were both differentiable and familiar.

Follow up tests for *Block* showed significant differences between each successive pair (all $p < 0.05$). The significant interaction between *Shape Distinctiveness* and *Block* can be seen in Figure 6.2, where the learning curve for FamiliarShape flattens before the other conditions (because users reached expertise far earlier in this condition).

Our second planned comparison (H2) investigates the effect of colour distinctiveness in icons that have no shape differentiability (Square vs. Square+Colour). A 2x4 RM-ANOVA (*Colour Distinctiveness X Block*) showed no effect of *Colour Distinctiveness* ($F_{1,19}$= 1.62, p= 0.2), and no interaction with *Block* ($F_{1,19}$= 0.56, p= 0.46).

Our fourth planned comparison (H4) looked at whether shape differentiability alone (with meaningless icons) would improve learning. We compared the Square and UnfamiliarShape conditions using a one-way RM-ANOVA, but found no difference ($F_{1,19}$= 0.27, p= 0.61).

6.2.2 Hovers

Similar to Study 1, the results for mean hover count in Study 2 closely mirror the completion time results. RM-ANOVA (*Shape Distinctiveness x Block*) showed effects of *Shape Distinctiveness* ($F_{1,19}$= 107.02, $p < 0.0001$, h^2= 0.73), *Block* ($F_{1,19}$= 290.45, $p < 0.0001$, h^2= 0.84) as well as an interaction between the two factors ($F_{1,19}$= 24.4, $p < 0.0001$, h^2= 0.18) on hover count (H1). Follow-up tests for *Block* showed significant differences (all $p < 0.05$) between each successive pair.

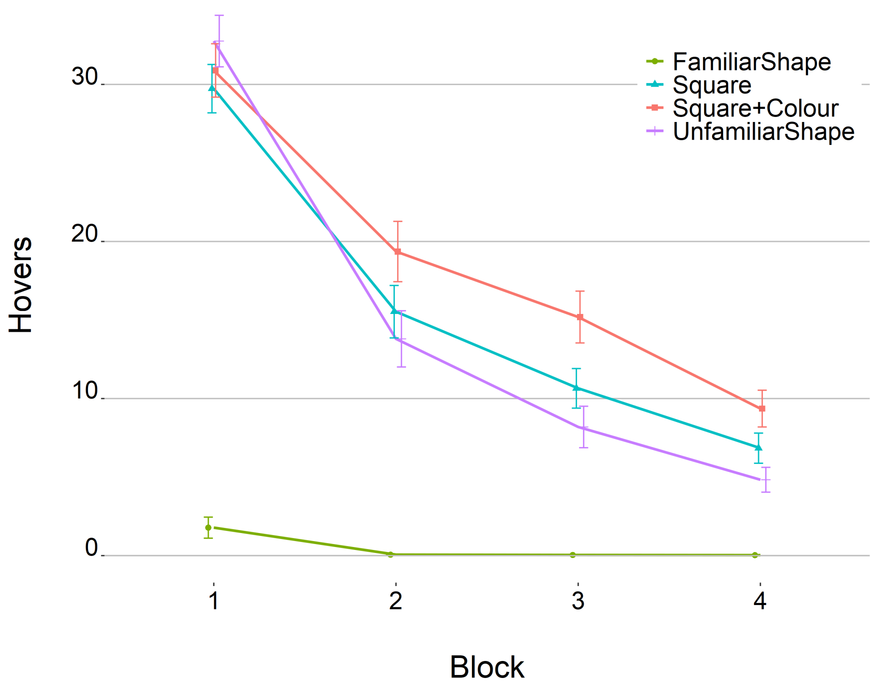


Figure 6.3: Mean hover count per block, by interface (s.e.).

As with the completion time results, the effect of Shape Distinctiveness appears to be largely due to the substantial effect of familiarity: in our third planned comparison, a one-way RM-ANOVA also showed a significant effect between UnfamiliarShape and FamiliarShape ($F_{1,19}$= 102.17, $p < 0.0001$, h^2= 0.73). T-tests also showed no significant difference between UnfamiliarShape and Square (p > 0.1), but showed that FamiliarShape was significantly different from all three other interfaces (all p < 0.001). Follow-up tests for *Block* showed significant differences between every successive pair except blocks 3 and 4.

In our second planned comparison (H2), a 2x4 RM-ANOVA found no effect of *Colour Distinctiveness* ($p > 0.15$) and no interaction with *Block* ($F_{1,19}$= 0.15, p= 0.71, h^2= 0.002).

In our fourth planned comparison (H4), a one-way RM-ANOVA found no effect of shape differentiability ($F_{1,19}$= 0.27, p= 0.61, h^2= 0.006).

6.2.3 Errors

We measured errors as the number of incorrect clicks before choosing the correct item. Data from one participant (who clicked instead of hovered) was removed. For all other participants, errors were very low, with an overall average of 0.037 errors per trial. RM-ANOVA showed no main effect of any of our main factors on errors (*Shape Distinctiveness*: $F_{1,19}$= 1.42, p= 0.24; *Block*: $F_{1,19}$= 0.39, p= 0.75; *Colour*: $F_{1,19}$= 1.67, p= 0.21; or *Familiarity*: $F_{1,19}$= 2.47, p= 0.13).

6.2.4 Subjective Responses and Comments

NASA-TLX responses were analyzed after performing an Aligned Rank Transformation [196]. Data from two participants, which was incomplete, was removed. The mean effort scores shown in Figure 6.4 mirror the trend in the performance data, in which FamiliarShape outperformed others in all measures. RM-ANOVA showed significant effects for all subjective measures. Follow-up tests showed significant (all $p < 0.05$) differences between FamiliarShape and every other condition in mental effort, perceived success, effort, and annoyance. Pearson's Chi-squared tests revealed significance for Easiest, Fastest, Fewest Errors, and Preference (X^2= 52.4, df= 3, p= < 0.0001). Overall, the FamiliarShape icons were greatly preferred—results are summarized in Table 6.1.

Participants' comments again echoed the performance results. Three participants stated that the uniformity in the Square condition was challenging; one said, *"it was really hard since everything looked the same."* Participants mostly remembered target locations, however two participants reported using row and column numbers to facilitate learning. Four participants reported using no strategies to help them remember the locations of the targets.

Four participants noted difficulties when attempting to use the colour information. For example, one participant stated *"I tried to use colour [in Square+Colour] but it didn't work super well."* Another remarked that they *"couldn't think of a consistent pattern to follow with colour/word association that was consistent for all."* The inability to assign the colours proved a hindrance to a participant; *"If I could assign it myself maybe I would have but having it assigned for me sucked. I can't associate that way. So it [Square+Colour] was the hardest one."* Thirteen participants reported memory tricks either independent of or as a compliment to colour such as *"trying to remember the general location"*, *"remembering their place by right, left, symmetry"*, and *"row number as a guide"*. Eight participants reported using

	Easiest	Fastest	Fewest Errors	Preference
Square	1	0	0	1
Square+Colour	0	1	0	0
UnfamiliarShape	0	0	1	0
FamiliarShape	19	19	19	19

Table 6.1: Summary of Study 2 preference survey results.

colour to facilitate learning the location of target icons in strategies; *"After finding the items the first time around I just tried to memorize what colour they were then when they'd come up again I'd only look at the boxes of that colour."*

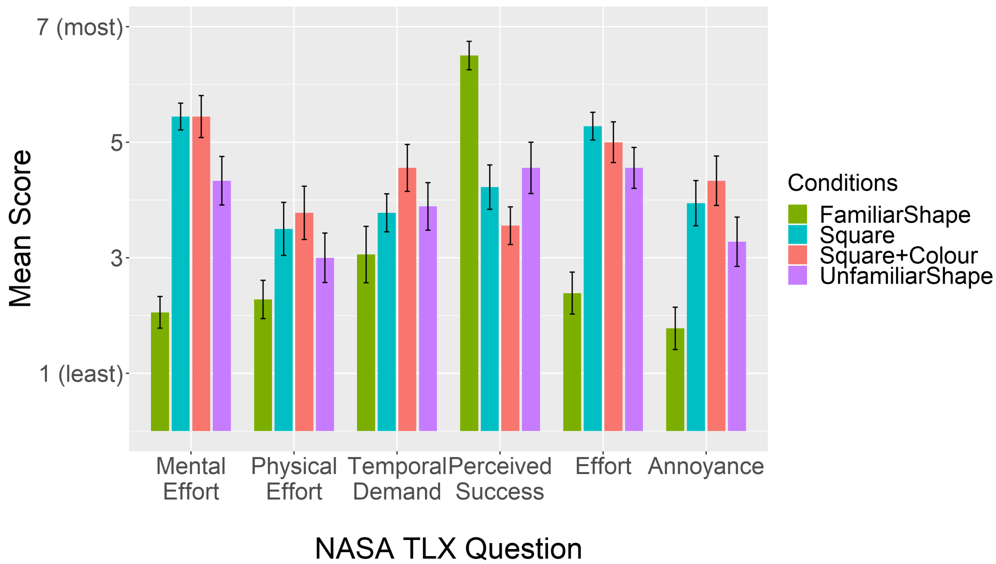


Figure 6.4: Mean NASA-TLX questions responses for Study 2 (s.e.).

The realistic representation of targets in FamiliarShape was found to be beneficial to eight participants (e.g., *"remembering the picture of each object, and my brain just brought me to where it was"*). Participants mostly did not use any strategies to help learn target icon locations. As the icons were *"easy to remember"*, participants did *"fast scan to recognize the shapes."*

Six participants stated that the distinct shapes of the UnfamiliarShape condition provided a connection that helped them to remember targets: for example, one participant reported that one target's icon *"looked like a bent cross"*, making it easier to remember the location. Participants made mental connections such as *"volume's shape kind of looks like an upside down V, so that's how I remembered that"* or *"I turned these shapes into something like I can remember eg. Contacts has a sideways C on top"*. Many participants remembered by utilising where the icons were located. A sampling of these comments include *"I just tried to memorize symbols I would pass over frequently"*, *"I tried to remember where the words were located based on like what words were around that word"*, *"knowing the column numbers where we have icons of interest"*, *"repeat where each one was after each additional symbol was added"*, and *"remembering their place and shapes by repeating"*.

7 Discussion and Conclusion

Our two studies provided the following findings:

- Colour distinctiveness did not improve learning in either study.
- Meaning was more valuable than visual distinctiveness.
- Adding multiple distinctive variables (colour and shape) also did not improve learning.
- Shape distinctiveness when coupled with meaning substantially improved learning, but shape distinctiveness on its own was not effective.
- Participant strategies suggested that they primarily try to search by meaning rather than visual characteristics.

In the following subsections, we consider explanations for these main results, limitations to our findings, and directions for future research.

7.1 Colour Distinctiveness Does Not Improve Icon Learning

Colour distinctiveness did not reduce completion time or hover count in either study, even when it was the only visual variable available (see Chapter 6.2.1 for results). One possible reason for this finding is that many participants did not use the colour cues, and instead only searched by meaning and spatial location—participants often reported creating and connecting stories to icons to remember them or memorizing by location rather than using colour as a visual landmark (see Chapter 5.2.4 and 6.2.4 for participant comments on their strategies for remembering where icons were). Another possible reason for colour's ineffectiveness may be that the colour cues interfered with one another, reducing the value of colour as a landmark. That is, because all icons were coloured, remembering only that an icon was "beside the blue one" did not uniquely identify a target (because there were several blue icons) [73, 82]. It is perhaps due to this reason that colour performed similarly to Square (the interface that repeated the colour grey across all sixty icons) in Study 2. By assigning one of a set number of colours to each icon within an interface, rather than creating landmarks within each icon, it renders the icon no different from one another since every icon possesses a colour. Furthermore, results from Study 1 and 2 are in agreement with Li et al. [101] that found visualizations containing more colours were not more memorable than single hue visualizations.

It is, however, possible that if there were fewer colour variations in the icons, colour might be a more effective landmark—in studies of artificial landmarks, for example, having fewer landmarks that were the same colour significantly improved performance [184] in a similar selection task. This also accords with observations expressed by

Horowitz et al. [143] that "congested" displays have little room for further visual variables to clamor for attention. It is also possible that colour interfered with participants' ability to see differences in the abstract shapes used in Study 1; that is, the colours used in the Abstract+Colour condition had poor contrast and thus may have reduced any potential effects of shape distinctiveness [93, 129].

7.2 Shape Distinctiveness was Only Effective with Meaning

When icons had a contextual level of meaning (see Chapter 3.1 for the detailed framework), we observed that participants would visually search using meaning as a memory cue; and when meaning was available, participants tended to disregard the landmarks created by differences in the icons' visual presentation. In icons with meaningless imagery, participants needed to rely more on absolute spatial memory—and without pre-existing knowledge of the icon mappings, participants had to find a prompted icon by laborious visual search (hovering one by one). Our findings confirm previous guidance about designing icons with clear meaning to help user navigation of an interface (e.g., [93, 53, 14, 76, 145, 78]), although our results extend this guidance to the value of meaning for longer-term learning of an interface as well. Moreover, our results also solidifies the understanding that shape distinctiveness is not as valuable as meaningfulness.

Furthermore, Study 2 showed that shape distinctiveness without meaning did not improve learning, and we speculate the reason for this condition's poor performance is similar to that of the colour conditions: namely, interference between similar-looking shapes may have prevented a shape's differentiability from being useful as a landmark. That is, because within visually consistent interfaces (i.e. four to five stroke Kanji characters sharing similar line widths in Study 2), remembering characteristics such as an icon had a "left facing stroke" did not uniquely identify a target because there were several icons with left facing strokes. As with colour, shape may still be useful as a landmark if there were fewer shapes that have more noticeable differences.

7.3 Design Implications and Generalizing the Results

Our results suggest that user learning of an interface is not hindered by the lack of visual distinctiveness in "flat" and subtle icon designs, and also clearly show the value of using concrete and familiar imagery. Therefore, designers can use flat and subtle icon styles without compromising learnability, as long as meaning is clearly conveyed. However, our study participants skewed younger (mean of 26 in study 1, and 24 in the pre-study and study 2) and as such, it remains to be seen if this finding extends to older audiences. We note that there are other potential factors in the use of flat icons that should be considered in addition to learning (e.g., whether users can tell that an on-screen object is in fact a clickable icon). Our results also raise the question of what designers should do in situations where they must create icons for commands or concepts that do not have obvious visual representations. The frequency with which we saw the "memory hook" strategy in our studies (i.e., looking for a connection between the image of the icon and the associated command) suggests that concrete imagery—even if not a direct representation of the underlying concept—may enable learning better than simply using distinctive visual variables. After all, images need to be recognized in some form before they can be remembered [12]. As suggested above, however, it may be that the value of colour or shape distinctiveness as landmarks could be improved, a topic we will consider in future studies.

While landmarks have long been proven to aid in navigating user interfaces [6, 181, 182], we theorize that having too many landmarks negated the efficacy of them. For example, if every icon was coloured, colour no longer renders an icon that possesses it unique from the other icons in the interface. As a participant noted, *"the colors or icons don't mean very much."* Furthermore, having many landmarks proved confusing to some participants. A participant remarked that colour *"really messed me up, I couldn't think of a consistent pattern to follow with colour/word association that was consistent for all."* This was also reflected by another participant when completing the UnfamiliarShape condition in Study 2: *"the symbols made no sense and were distracting from trying to find the words and made it worse then just having grey boxes...I had to check all the symbols for the words nearly every time."* However, despite our results showing colour providing no benefit to the learnability of a GUI, many participants reported using colour to help remember where target icons were (see chapters 4.2.4 and 6.2.4 for detailed participant comments). As such, an implication for consideration to designers is that the usage of colour can still be useful as a tool for users to orient themselves within an interface.

The presentation of the study interfaces we evaluated can be observed in interfaces for various desktop productivity applications—our interfaces were modeled closely after Microsoft Word's toolbar. Specifically, like Word, our study interfaces grouped commands in a tab-based ribbon menu located at the top of an interface. Word and our study interfaces both present a multitude of commands at a time which may overwhelm new or returning users when learning commands. Similarly to Word, our icons did not display icon names in the icon shape. However, unlike our study interfaces, Word employs semantic grouping to emphasize similar commands and differences in size to anchor users. Arguably, these form landmarks that may better orient users when navigating their ribbon menu. Furthermore, commercial productivity software programs are expected to be used by a wide range of ages, whereas our participant pool primarily consisted of college-aged students. However, we expect the challenges observed in learning the various interfaces evaluated in this study to be consistent across age groups and commercial interfaces; unfamiliar shapes will remain unfamiliar unless a user creates a mnemonic connection to the shape and that icons without meaning will be harder to navigate than meaningful icons. As such, we generally expect our results to translate to commercial interfaces.

7.4 Limitations and Future Work

There are several ways in which our studies could not exactly replicate various factors of real-world interface learning, and these suggest possibilities for future research. A limitation in our studies was the short time available for learning—users typically learn an interface in a much slower fashion, and in the context of real tasks. However, in the context of our study, participants had neither tasks to associate the target icons with nor the time to reach ultimate performance within an interface.

In addition, we tested only immediate recall, not retention after a time period, and we did not test transfer from the training task back to a real-world task with the interface. While common visual variables such as shape and colour did not reflect meaningful gains in our study, it is worth investigating the retention of the target items after a period of time. Moreover, further studies should consider whether colour and shape may prove useful in later phases of the intermodal performance curve (see chapter 2.3.2). Specifically, further studies should consider whether visual variables hold any level of value for users in the extended learnability and ultimate performance phases. For example, if users need to select an icon they know the general location of having gained familiarity with a layout, do they use the icon's individual visual variables as a way to quickly identify them within the immediate area? Do they still remember the visual variables characteristics of the icon they're looking for? Would meaning be more valuable than colour when users remember where target icons are? We plan retention and transfer phases in our future studies.

Further studies should also consider whether colour or shape differentiability could be more effective if there are fewer items in the set that are different, thus providing a better anchor for spatial learning. One implementation would involve strategically placed icons that are designed to catch the user's attention (using colour or shape) within the toolbar; these icons could anchor memory and potentially improve learning. Salience can also be explored through unique usages of context. For example, a possible avenue of exploration involves employing select meaningfully different icons in a manner similar to the above, sparingly, to serve as anchors for spatial learning. As our results demonstrated, meaning is powerful in learning icons.

7.5 Conclusion

Icons are a ubiquitous mechanism for representing commands in an interface [14], and learning the icons in an interface is a major part of becoming an expert with that system. Despite the prevalence of icons, toolbars, and ribbons, however, little is known about the effects of icon design on learnability. We carried out two studies to test whether differentiability in two visual variables—colour and shape—would improve learning of icons in a 60-item toolbar. Our results showed that our manipulations of these variables did not have significant effects on learning or performance, and that the concreteness and meaning of the icon's imagery was far more effective in helping users learn and recall targets. Our studies provide new empirical evidence for existing guidelines that suggest an icon that is contextual or familiar will be more learnable and easier to navigate. This work increases understanding of how users learn new icons and the relative roles that visual variables and cognitive factors play in users' spatial learning and expertise development.

REFERENCES

[1] Free icons, clipart illustrations, photos, and music.

[2] Icons = google fonts.

[3] ADOBE. Adobe photoshop, 2020.

[4] ALEXANDER, J. Icons web development website design flat, 2017. https://pixabay.com/illustrations/icons-web-development-website-design-2188729/.

[5] ALEXANDER, J., COCKBURN, A., FITCHETT, S., GUTWIN, C., AND GREENBERG, S. Revisiting read wear: analysis, design, and evaluation of a footprints scrollbar. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2009), pp. 1665–1674.

[6] ALEXANDER, J., COCKBURN, A., FITCHETT, S., GUTWIN, C., AND GREENBERG, S. Revisiting read wear: analysis, design, and evaluation of a footprints scrollbar. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems - CHI '09 (New York, NY, USA, 2009), ACM, pp. 1665–1674.

[7] ALVAREZ, G. A., AND CAVANAGH, P. The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychological science 15, 2 (2004), 106–111.

[8] ANDERSON, J. R. Learning and memory: An integrated approach. John Wiley & Sons Inc, 2000.

[9] ANDERSON, J. R. Cognitive psychology and its implications. Macmillan, 2005.

[10] ARTHUR, P., AND PASSINI, R. Wayfinding: people, signs, and architecture. Focus Strategic Communications, 1992.

[11] BACKHAUS, N., TRAPP, A. K., AND THÜRING, M. Skeuomorph versus flat design: User experience and age-related preferences. In International Conference of Design, User Experience, and Usability (2018), Springer, pp. 527–542.

[12] BAINBRIDGE, W. A., DILKS, D. D., AND OLIVA, A. Memorability: A stimulus-driven perceptual neural signature distinctive from memory. NeuroImage 149 (2017), 141–152.

[13] BAR, M., AND NETA, M. Humans prefer curved visual objects. Psychological science 17, 8 (2006), 645–648.

[14] BARR, P., NOBLE, J., AND BIDDLE, R. Icons r icons. In Proceedings of the Fourth Australasian User Interface Conference on User Interfaces 2003 - Volume 18 (AUS, 2003), AUIC '03, Australian Computer Society, Inc., p. 25–32.

[15] BARRETT, R., KANDOGAN, E., MAGLIO, P. P., HABER, E. M., TAKAYAMA, L. A., AND PRABAKER, M. Field studies of computer system administrators: analysis of system management tools and practices. In Proceedings of the 2004 ACM conference on Computer supported cooperative work (2004), pp. 388–395.

[16] BATEMAN, S., MANDRYK, R. L., GUTWIN, C., GENEST, A., MCDINE, D., AND BROOKS, C. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In Proceedings of the SIGCHI conference on human factors in computing systems (2010), pp. 2573–2582.

[17] BERNSTEIN, D. Essentials of psychology. Cengage Learning, 2018.

[18] BHAVNANI, S. K., AND JOHN, B. E. The strategic use of complex computer systems. Human–computer interaction 15, 2-3 (2000), 107–137.

[19] BLATTNER, M. M., SUMIKAWA, D. A., AND GREENBERG, R. M. Earcons and icons: Their structure and common design principles. Hum.-Comput. Interact. 4, 1 (Mar. 1989), 11–44.

[20] BLAYNEE, J., KREPS, D., KUTAR, M., AND GRIFFITHS, M. Collaborative hci and ux: longitudinal diary studies as a means of uncovering barriers to digital adoption. In Proceedings of the 30th International BCS Human Computer Interaction Conference 30 (2016), pp. 1–6.

[21] BRODAL, P. The central nervous system: structure and function. Oxford University Press, 2004.

[22] BURLINSON, D., SUBRAMANIAN, K., AND GOOLKASIAN, P. Open vs. closed shapes: New perceptual categories? IEEE transactions on visualization and computer graphics 24, 1 (2017), 574–583.

[23] BURMISTROV, I., ZLOKAZOVA, T., IZMALKOVA, A., AND LEONOVA, A. Flat design vs traditional design: Comparative experimental study. In IFIP Conference on Human-Computer Interaction (2015), Springer, pp. 106–114.

[24] CARD, S. K., MORAN, T. P., AND NEWELL, A. The psychology of human-computer interaction. Crc Press, 2018.

[25] CARRASCO, M. Visual attention: The past 25 years. Vision research 51, 13 (2011), 1484–1525.

[26] CARRASCO, M., EVERT, D. L., CHANG, I., AND KATZ, S. M. The eccentricity effect: Target eccentricity affects performance on conjunction searches. Perception & psychophysics 57, 8 (1995), 1241–1261.

[27] CARROLL, J. M., AND ROSSON, M. B. Paradox of the active user. In Interfacing thought: Cognitive aspects of human-computer interaction. 1987, pp. 80–111.

[28] CARTER, R. Visual search with color. Journal of experimental psychology. Human perception and performance 8, 1 (February 1982), 127—136.

[29] CHANWIMALUENG, W., AND RAPEEPISARN, K. A study of the recognitions and preferences on abstract and concrete icon styles on smart phone from easterners and westerners' point of view. Proceedings - International Conference on Machine Learning and Cybernetics 4 (07 2013), 1613–1619.

[30] CHO, M., KWON, S., NA, N., SUK, H.-J., AND LEE, K. The elders preference for skeuomorphism as app icon style. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (2015), pp. 899–904.

[31] CHRIST, R. E. Review and analysis of color coding research for visual displays. Human Factors 17, 6 (1975), 542–570.

[32] COCKBURN, A., GUTWIN, C., AND ALEXANDER, J. Faster document navigation with space-filling thumbnails. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2006), CHI '06, ACM, p. 1–10.

[33] COCKBURN, A., GUTWIN, C., AND GREENBERG, S. A predictive model of menu performance. In Proceedings of the SIGCHI conference on Human factors in computing systems (2007), pp. 627–636.

[34] COCKBURN, A., GUTWIN, C., SCARR, J., AND MALACRIA, S. Supporting novice to expert transitions in user interfaces. ACM Comput. Surv. 47, 2 (Nov. 2014).

[35] COCKBURN, A., KRISTENSSON, P. O., ALEXANDER, J., AND ZHAI, S. Hard lessons: Effort-inducing interfaces benefit spatial learning. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2007), CHI '07, ACM, p. 1571–1580.

[36] COHEN, J. Eta-squared and partial eta-squared in communication science. Human Communication Research 28, 56 (1973), 473–490.

[37] CONNOR, C. E., EGETH, H. E., AND YANTIS, S. Visual attention: bottom-up versus top-down. Current biology 14, 19 (2004), R850–R852.

[38] CONWAY, B. R., AND LIVINGSTONE, M. S. Spatial and temporal properties of cone signals in alert macaque primary visual cortex. Journal of Neuroscience 26, 42 (2006), 10826–10846.

[39] COWAN, N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and brain sciences 24, 1 (2001), 87–114.

[40] CREAGER, J. H., AND GILLAN, D. J. Toward understanding the findability and discoverability of shading gradients in almost-flat design. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2016), vol. 60, SAGE Publications Sage CA: Los Angeles, CA, pp. 339–343.

[41] CUI, A. Outline icons (free psd), 2014. Available at https://www.behance.net/gallery/18385435/Outline-Icons-(Free-PSD.

[42] CZERWINSKI, M., HORVITZ, E., AND CUTRELL, E. Subjective duration assessment: An implicit probe for software usability. In Proceedings of IHM-HCI 2001 conference (2001), vol. 2, pp. 167–170.

[43] CZERWINSKI, M., VAN DANTZICH, M., ROBERTSON, G. G., AND HOFFMAN, H. G. The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3d. In INTERACT (1999), pp. 163–170.

[44] DARKEN, R. P., AND SIBERT, J. L. Wayfinding strategies and behaviors in large virtual worlds. In Proceedings of the SIGCHI conference on Human factors in computing systems (1996), pp. 142–149.

[45] DEEGAN, P. Lame icons in office 365 isn't helped by adding 300 more, 2019.

[46] DEYOE, E. A., AND VAN ESSEN, D. C. Segregation of efferent connections and receptive field properties in visual area v2 of the macaque. Nature 317, 6032 (1985), 58–61.

[47] DUNCAN, J., AND HUMPHREYS, G. W. Visual search and stimulus similarity. Psychological review 96, 3 (1989), 433.

[48] EGETH, H. Attention and preattention. In Psychology of Learning and Motivation, vol. 11. Elsevier, 1977, pp. 277–320.

[49] EGETH, H. E., VIRZI, R. A., AND GARBART, H. Searching for conjunctively defined targets. Journal of Experimental Psychology: Human Perception and Performance 10, 1 (1984), 32.

[50] EHRET, B. D. Learning where to look: Location learning in graphical user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2002), CHI '02, ACM, p. 211–218.

[51] EYSENCK, M. Attention and arousal: Cognition and performance. Springer Science & Business Media, 2012.

[52] FADZIL, M. I., OMAR, M. W., AND MURAD, M. H. The power of color on packaging. In 2011 IEEE Colloquium on Humanities, Science and Engineering (2011), IEEE, pp. 761–764.

[53] FERREIRA, J., NOBLE, J., AND BIDDLE, R. A case for iconic icons. In Proceedings of the 7th Australasian User interface conference-Volume 50 (2006), Australian Computer Society, Inc., pp. 97–100.

[54] FERRERA, V. P., NEALEY, T. A., AND MAUNSELL, J. H. Mixed parvocellular and magnocellular geniculate signals in visual area v4. Nature 358, 6389 (1992), 756–758.

[55] FITTS, P. M., AND POSNER, M. I. Human performance. Brooks/Cole, 1967.

[56] GALASHAN, D., AND SIEMANN, J. Differences and similarities for spatial and feature-based selective attentional orienting. Frontiers in neuroscience 11 (2017), 283.

[57] GALE, A., CARR, K., AND BROGAN, D. Visual Search 2: Proceedings Of The 2nd International Conference On Visual Search. CRC Press, 1993.

[58] GAO, B., KIM, B., KIM, J.-I., AND KIM, H. Amphitheater layout with egocentric distance-based item sizing and landmarks for browsing in virtual reality. International Journal of Human–Computer Interaction 35, 10 (2019), 831–845.

[59] GAO, B., KIM, H., KIM, B., AND KIM, J.-I. Artificial landmarks to facilitate spatial learning and recalling for curved visual wall layout in virtual reality. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (2018), IEEE, pp. 475–482.

[60] GAUR, V., UDDIN, M. S., AND GUTWIN, C. Multiplexing spatial memory: increasing the capacity of fasttap menus with multiple tabs. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (2018), pp. 1–13.

[61] GITTINS, D. Icon-based human-computer interaction. International Journal of Man-Machine Studies 24, 6 (1986), 519–543.

[62] GOLDSTEIN, E. B. Cognitive psychology: Connecting mind, research and everyday experience. Nelson Education, 2014.

[63] GOODALE, M. A., AND MILNER, A. Separate visual pathways for perception and action. Trends in Neurosciences 15 (1992), 20 – 25.

[64] GREGORY, R. Perception. Zangwill, 1987.

[65] GROSS, S., BARDZELL, J., AND BARDZELL, S. Skeu the evolution: skeuomorphs, style, and the material of tangible interactions. In Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction (2014), pp. 53–60.

[66] GROSSMAN, T., FITZMAURICE, G., AND ATTAR, R. A survey of software learnability: metrics, methodologies and guidelines. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2009), pp. 649–658.

[67] GUTWIN, C., AND COCKBURN, A. Improving list revisitation with listmaps. In Proceedings of the working conference on Advanced visual interfaces (2006), pp. 396–403.

[68] GUTWIN, C., COCKBURN, A., AND LAFRENIERE, B. Testing the rehearsal hypothesis with two fasttap interfaces. In Proceedings of the 41st Graphics Interface Conference (2015), pp. 223–231.

[69] GUTWIN, C., COCKBURN, A., SCARR, J., MALACRIA, S., AND OLSON, S. C. Faster command selection on tablets with fasttap. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2014), pp. 2617–2626.

[70] HART, S. G., AND STAVELAND, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology 52 (1988), 139–183.

[71] HEALEY, C., AND ENNS, J. Attention and visual memory in visualization and computer graphics. IEEE transactions on visualization and computer graphics 18, 7 (2011), 1170–1188.

[72] HEGARTY, M., MONTELLO, D. R., RICHARDSON, A. E., ISHIKAWA, T., AND LOVELACE, K. Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. Intelligence 34, 2 (2006), 151–176.

[73] HICK, W. E. On the rate of gain of information. Quarterly Journal of Experimental Psychology 4, 1 (mar 1952), 11–26.

[74] HILLSTROM, A. P. Repetition effects in visual search. Perception & psychophysics 62, 4 (2000), 800–817.

[75] HINE, T. The total package: the secret history and hidden meanings of boxes, bottles, cans and other persuasive containers. Back Bay Books, 1997.

[76] HORTON, W. Designing icons and visual symbols. In Conference companion on Human factors in computing systems (1996), pp. 371–372.

[77] Hou, K.-C., Ho, C.-H., et al. A preliminary study on aesthetic of apps icon design. In IASDR 2013 5th International Congress of International Association of Societies of Design Research (2013), pp. 1–12.

[78] Huang, S.-C., Bias, R. G., and Schnyer, D. How are icons processed by the brain? neuroimaging measures of four types of visual stimuli used in information systems. Journal of the association for information science and technology 66, 4 (2015), 702–720.

[79] Hubel, D. H., and Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology 195, 1 (1968), 215–243.

[80] Huff, T., Mahabadi, N., and Tadi, P. Neuroanatomy, visual cortex. StatPearls [Internet] (2020).

[81] Humphreys, G. W., Quinlan, P. T., and Riddoch, M. J. Grouping processes in visual search: Effects with single-and combined-feature targets. Journal of Experimental Psychology: General 118, 3 (1989), 258.

[82] Hyman, R. Stimulus information as a determinant of reaction time. Journal of Experimental Psychology 45, 3 (1953), 188–196.

[83] Itti, L., and Koch, C. Computational modelling of visual attention. Nature reviews neuroscience 2, 3 (2001), 194–203.

[84] Itti, L., Koch, C., and Niebur, E. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence 20, 11 (1998), 1254–1259.

[85] Jiang, Y., Song, J.-H., and Rigas, A. High-capacity spatial contextual memory. Psychonomic bulletin & review 12, 3 (2005), 524–529.

[86] Johnson, M. A. Color vision in the peripheral retina. American journal of optometry and physiological optics 63, 2 (1986), 97–103.

[87] Kahopkins. How can we improve your desktop outlook 2016 experience?, 2016.

[88] Kandel, E. R., Schwartz, J. H., Jessell, T. M., of Biochemistry, D., Jessell, M. B. T., Siegelbaum, S., and Hudspeth, A. Principles of neural science, vol. 4. McGraw-hill New York, 2000.

[89] Kang, X. The effect of color on short-term memory in information visualization. In proceedings of the 9th International Symposium on Visual Information Communication and Interaction (2016), pp. 144–145.

[90] Kaptelinin, V. Item recognition in menu selection: the effect of practice. In INTERACT'93 and CHI'93 conference companion on Human factors in computing systems (1993), pp. 183–184.

[91] Kim, K., Erickson, A., Lambert, A., Bruder, G., and Welch, G. Effects of dark mode on visual fatigue and acuity in optical see-through head-mounted displays. In Symposium on Spatial User Interaction (2019), pp. 1–9.

[92] Koehn, C. Iconic icons: Designing the world of windows. Medium (2020).

[93] Kurniawan, S. H. A rule of thumb of icons' visual distinctiveness. In Proceedings on the 2000 conference on Universal Usability (2000), pp. 159–160.

[94] LaBerge, D. Automatic information processing: A review. Attention and performance IX (1981), 173–186.

[95] Labrecque, L. I., and Milne, G. R. Exciting red and competent blue: the importance of color in marketing. Journal of the Academy of Marketing Science 40, 5 (2012), 711–727.

[96] Lafreniere, B., Gutwin, C., Cockburn, A., and Grossman, T. Faster command selection on touchscreen watches. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (2016), pp. 4663–4674.

[97] Lakoff, G., and Johnson, M. Metaphors we live by. University of Chicago press, 2008.

[98] LANE, D. M., NAPIER, H. A., PERES, S. C., AND SANDOR, A. Hidden costs of graphical user interfaces: Failure to make the transition from menus and icon toolbars to keyboard shortcuts. International Journal of Human-Computer Interaction 18, 2 (2005), 133–144.

[99] LAPPIN, J. S., AND UTTAL, W. R. Does prior knowledge facilitate the detection of visual targets in random noise? Perception & Psychophysics 20, 5 (1976), 367–374.

[100] LI, C. F., SHI, H. T., HUANG, J. J., AND CHEN, L. Y. Two typical symbols in human-machine interactive interface. In Applied Mechanics and Materials (2014), vol. 635, Trans Tech Publ, pp. 1659–1665.

[101] LI, R., AND CHEN, J. Toward a deep understanding of what makes a scientific visualization memorable. In 2018 IEEE Scientific Visualization Conference (SciVis) (2018), IEEE, pp. 26–31.

[102] LIDWELL, W., HOLDEN, K., AND BUTLER, J. Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design. Rockport Pub, 2010.

[103] LINDGAARD, G., CHESSARI, J., AND IHSEN, E. Icons in telecommunications-what makes pictorial information comprehensible to the user. Australian telecommunication research 21, 2 (1987), 17–29.

[104] LINDSAY, P. H., AND NORMAN, D. A. Human information processing: An introduction to psychology. Academic press, 2013.

[105] LODDING, K. Iconic interfacing. IEEE Computer graphics and applications 2 (1983), 11–20.

[106] LYNCH, K. The image of the city, vol. 11. MIT press, 1960.

[107] MAGUIRE, E. A., BURGESS, N., AND O'KEEFE, J. Human spatial navigation: cognitive maps, sexual dimorphism, and neural substrates. Current opinion in neurobiology 9, 2 (1999), 171–177.

[108] MALJKOVIC, V., AND NAKAYAMA, K. Priming of pop-out: I. role of features. Memory & cognition 22, 6 (1994), 657–672.

[109] MAUNSELL, J. H., AND VAN ESSEN, D. C. Functional properties of neurons in middle temporal visual area of the macaque monkey. i. selectivity for stimulus direction, speed, and orientation. Journal of neurophysiology 49, 5 (1983), 1127–1147.

[110] MICROSOFT. Use of microsoft copyrighted content. https://www.microsoft.com/en-us/legal/intellectualproperty/copyright/permissions.

[111] MICROSOFT. Microsoft word, 2020.

[112] MILLER, E. K. The prefontral cortex and cognitive control. Nature reviews neuroscience 1, 1 (2000), 59–65.

[113] MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological review 63, 2 (1956), 81.

[114] MOLLASHAHI, E. S., UDDIN, M. S., AND GUTWIN, C. Improving revisitation in long documents with two-level artificial-landmark scrollbars. In Proceedings of the 2018 International Conference on Advanced Visual Interfaces (2018), pp. 1–9.

[115] MONAHAN, J. S., AND LOCKHEAD, G. R. Identification of integral stimuli. Journal of Experimental Psychology: General 106, 1 (1977), 94.

[116] MOON, J. M., AND FU, W.-T. Where is my stuff? augmenting finding and re-finding information by spatial locations and icon luminance. In International Conference on Foundations of Augmented Cognition (2009), Springer, pp. 58–67.

[117] MOYES, J., AND JORDAN, P. W. Icon design and its effect on guessability, learnability, and experienced user performance. People and Computers 8 (1993), 49–60.

[118] NAGY, A. L., AND SANCHEZ, R. R. Critical color differences determined with a visual search task. JOSA A 7, 7 (1990), 1209–1217.

[119] NEISSER, U. Visual search. Scientific American 210, 6 (1964), 94–103.

[120] NEISSER, U. Cognitive psychology: Classic edition. Psychology Press, 2014.

[121] NEVEU, M., AND JEFFERY, G. Chiasm formation in man is fundamentally different from that in the mouse. Eye 21, 10 (2007), 1264–1270.

[122] NEWELL, A., AND ROSENBLOOM, P. S. Mechanisms of skill acquisition and the law of practice. Cognitive skills and their acquisition 1, 1981 (1981), 1–55.

[123] NG, A. W., AND CHAN, A. H. What makes an icon effective? In AIP Conference Proceedings (2009), vol. 1089, American Institute of Physics, pp. 104–114.

[124] NIELSEN, J. Usability engineering: Morgan kaufmann publishers. Inc.–1993 (1993).

[125] ODELL, D. L., DAVIS, R. C., SMITH, A., AND WRIGHT, P. K. Toolglasses, marking menus, and hotkeys: a comparison of one and two-handed command selection techniques.

[126] OKAWA, H., AND SAMPATH, A. P. Optimization of single-photon response transmission at the rod-to-rod bipolar synapse. Physiology 22, 4 (2007), 279–286.

[127] ORBAN, G., KENNEDY, H., AND BULLIER, J. Velocity sensitivity and direction selectivity of neurons in areas v1 and v2 of the monkey: influence of eccentricity. Journal of Neurophysiology 56, 2 (1986), 462–480.

[128] OSWALD, D., KOLB, S., ET AL. Flat design vs. skeuomorphism–effects on learnability and image attributions in digital product interfaces. In DS 78: Proceedings of the 16th International conference on Engineering and Product Design Education (E&PDE14), Design Education and Human Technology Relations, University of Twente, The Netherlands, 04-05.09. 2014 (2014), pp. 402–407.

[129] PAL, R., MUKHERJEE, J., AND MITRA, P. How do warm colors affect visual attention? In Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing (2012), pp. 1–8.

[130] PARASURAMAN, R. Vigilance, monitoring, and search. John Wiley & Sons, 1986.

[131] PASUPATHY, A., AND CONNOR, C. E. Shape representation in area v4: position-specific tuning for boundary conformation. Journal of neurophysiology 86, 5 (2001), 2505–2519.

[132] PEIRCE, C. S. Collected writings (8 vols.), 1931.

[133] PERRY, C. J., AND FALLAH, M. Feature integration and object representations along the dorsal stream visual hierarchy. Frontiers in computational neuroscience 8 (2014), 84.

[134] PHINNEY, R. E., AND SIEGEL, R. M. Speed selectivity for optic flow in area 7a of the behaving macaque. Cerebral Cortex 10, 4 (2000), 413–421.

[135] POSTMA, A., AND DE HAAN, E. H. What was where? memory for object locations. The Quarterly Journal of Experimental Psychology Section A 49, 1 (1996), 178–199.

[136] PRENSKY, M. Digital natives, digital immigrants part 1. On the horizon (2001).

[137] PURCHASE, H. Defining multimedia. IEEE multimedia 5, 1 (1998), 8–15.

[138] PURVES, D., AUGUSTINE, G., FITZPATRICK, D., KATZ, L., LAMANTIA, A., MCNAMARA, J., AND WILLIAMS, S. Neuroscience 2nd edition. sunderland (ma) sinauer associates. Inc, 2001.

[139] RAMEL, D. Developers revolt against new vs code icons, 2017.

[140] RIMBERT, S. Jacques bertin, sémiologie graphique: les diagrammes, les réseaux, les cartes. In Annales de géographie (1975), vol. 84, Persée-Portail des revues scientifiques en SHS, pp. 241–242.

[141] ROGERS, Y. Icons at the interface: their usefulness. Interacting with computers 1, 1 (1989), 105–117.

[142] ROORDA, A., AND WILLIAMS, D. R. The arrangement of the three cone classes in the living human eye. Nature 397, 6719 (1999), 520–522.

[143] ROSENHOLTZ, R., LI, Y., MANSFIELD, J., AND JIN, Z. Feature congestion: a measure of display clutter. In Proceedings of the SIGCHI conference on Human factors in computing systems (2005), pp. 761–770.

[144] RUSSELL-GOUGH, B. Normalized response spectra of human cones, to monochromatic spectral stimuli, with wavelength given in nanometers, 2009. https://en.wikipedia.org/wiki/Color_vision#/media/File:Cone-fundamentals-with-srgb-spectrum.svg.

[145] SATCHAROEN, K. Icon concreteness effect on selection speed and accuracy. In Proceedings of the 2018 10th International Conference on Computer and Automation Engineering (2018), pp. 107–110.

[146] SCARR, J. Understanding and Exploiting Spatial Memory in the Design of Efficient Command Selection Interfaces. PhD thesis, University of Canterbury. Computer Science and Software Engineering, 2014.

[147] SCARR, J., COCKBURN, A., GUTWIN, C., AND BUNT, A. Improving command selection with commandmaps. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2012), pp. 257–266.

[148] SCARR, J., COCKBURN, A., GUTWIN, C., BUNT, A., AND CECHANOWICZ, J. E. The usability of commandmaps in realistic tasks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2014), pp. 2241–2250.

[149] SCARR, J., COCKBURN, A., GUTWIN, C., AND MALACRIA, S. Testing the robustness and performance of spatially consistent interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2013), pp. 3139–3148.

[150] SCARR, J., COCKBURN, A., GUTWIN, C., AND QUINN, P. Dips and ceilings: understanding and supporting transitions to expertise in user interfaces. In Proceedings of the sigchi conference on human factors in computing systems (2011), pp. 2741–2750.

[151] SCARR, J., COCKBURN, A., GUTWIN, C., AND QUINN, P. Dips and ceilings: understanding and supporting transitions to expertise in user interfaces. In Proceedings of the sigchi conference on human factors in computing systems (2011), pp. 2741–2750.

[152] SCARR, J. L. Understanding and exploiting spatial memory in the design of efficient command selection interfaces.

[153] SCHACTER, D., GILBERT, D. T., AND WEGNER, D. M. Psychology (2nd Edition). Worth, New York, 2011.

[154] SCHNEIDERMEIER, T., HERTLEIN, F., AND WOLFF, C. Changing paradigm–changing experience? In International Conference of Design, User Experience, and Usability (2014), Springer, pp. 371–382.

[155] SCHRAMM, K., GUTWIN, C., AND COCKBURN, A. Supporting Transitions to Expertise in Hidden Toolbars. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2016), ACM, pp. 4687–4698.

[156] SHELTON, A. L., AND MCNAMARA, T. P. Systems of spatial reference in human memory. Cognitive psychology 43, 4 (2001), 274–310.

[157] SIEGEL, A. W., AND WHITE, S. H. The development of spatial representations of large-scale environments. In Advances in child development and behavior, vol. 10. Elsevier, 1975, pp. 9–55.

[158] SILVENNOINEN, J. M., AND JOKINEN, J. P. Aesthetic appeal and visual usability in four icon design eras. In Proceedings of the Conference on Human Factors in Computing Systems (2016), pp. 4390–4400.

[159] SILVIA, P. J., AND BARONA, C. M. Do people prefer curved objects? angularity, expertise, and aesthetic preference. Empirical studies of the arts 27, 1 (2009), 25–42.

[160] SIMON, H. A. Theories of decision-making in economics and behavioural science. In Surveys of economic theory. Springer, 1966, pp. 1–28.

[161] SMART, S., AND SZAFIR, D. A. Measuring the separability of shape, size, and color in scatterplots. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (2019), pp. 1–14.

[162] SMITH, S. L. Color coding and visual search. Journal of Experimental Psychology 64, 5 (1962), 434.

[163] SMITH, S. L., AND THOMAS, D. W. Color versus shape coding in information displays. Journal of Applied Psychology 48, 3 (1964), 137.

[164] SOLOMON, S. G., PEIRCE, J. W., AND LENNIE, P. The impact of suppressive surrounds on chromatic properties of cortical neurons. Journal of Neuroscience 24, 1 (2004), 148–160.

[165] SOSA-TZEC, O. Design tensions: interaction criticism on instagram's mobile interface. In Proceedings of the 37th ACM International Conference on the Design of Communication (2019), pp. 1–10.

[166] SOWWARDS, A. A collection of minimal icons for subtle design work., 2014.

[167] SPILIOTOPOULOS, K., RIGOU, M., AND SIRMAKESSIS, S. A comparative study of skeuomorphic and flat design from a ux perspective. Multimodal Technologies and Interaction 2, 2 (2018), 31.

[168] STARR, C., EVERS, C., AND STARR, L. Biology: concepts and applications. Nelson Education, 2010.

[169] STICKEL, C., POHL, H.-M., AND MILDE, J.-T. Cutting edge design or a beginner's mistake?–a semiotic inspection of ios7 icon design changes. In International Conference of Design, User Experience, and Usability (2014), Springer, pp. 358–369.

[170] SUMMERFIELD, C., AND EGNER, T. Feature-based attention and feature-based expectation. Trends in cognitive sciences 20, 6 (2016), 401–404.

[171] SWAIN, M. J., AND BALLARD, D. H. Color indexing. International journal of computer vision 7, 1 (1991), 11–32.

[172] SZMAJDA, B. A., GRÜNERT, U., AND MARTIN, P. R. Retinal ganglion cell inputs to the koniocellular pathway. Journal of Comparative Neurology 510, 3 (2008), 251–268.

[173] THE GIMP DEVELOPMENT TEAM. Gimp, 2020.

[174] THEEUWES, J. Feature-based attention: It is all bottom-up priming. Philosophical Transactions of the Royal Society B: Biological Sciences 368, 1628 (2013), 20130055.

[175] THORNDYKE, P. W., AND GOLDIN, S. E. Spatial learning and reasoning skill. In Spatial orientation. Springer, 1983, pp. 195–217.

[176] TILLMAN, B. Creative Character Design. Focal Press, Oxford, England, July 2011.

[177] TIMFAO. Office 365 icons changed, 2018.

[178] TREISMAN, A. Focused attention in the perception and retrieval of multidimensional stimuli. Perception & Psychophysics 22, 1 (1977), 1–11.

[179] TREISMAN, A. M., AND GELADE, G. A feature-integration theory of attention. Cognitive psychology 12, 1 (1980), 97–136.

[180] TSAL, Y., AND LAMY, D. Attending to an object's color entails attending to its location: Support for location-special views of visual attention. Perception & psychophysics 62, 5 (2000), 960–968.

[181] UDDIN, M. S. Improving Multi-Touch Interactions Using Hands as Landmarks. PhD thesis, University of Saskatchewan, 2016.

[182] UDDIN, M. S. Use of Landmarks to Design Large and Efficient Command Interfaces. In Proceedings of the ACM Companion on Interactive Surfaces and Spaces (New York, NY, USA, 2016), ACM, pp. 13–17.

[183] UDDIN, M. S., AND GUTWIN, C. 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 (2016), pp. 205–214.

[184] UDDIN, M. S., GUTWIN, C., AND COCKBURN, A. 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 (2017), pp. 3843–3855.

[185] UDDIN, M. S., GUTWIN, C., AND GOGUEY, A. Using artificial landmarks to improve revisitation performance and spatial learning in linear control widgets. In Proceedings of ACM symposium on Spatial User Interaction - SUI 2017 (Brighton, United Kingdom, 2017), ACM, pp. 48–57.

[186] UDDIN, M. S., GUTWIN, C., AND LAFRENIERE, B. 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 (2016), pp. 5836–5848.

[187] URBANO, I. C. V. P., GUERREIRO, J. P. V., AND NICOLAU, H. M. A. A. From skeuomorphism to flat design: age-related differences in performance and aesthetic perceptions. Behaviour & Information Technology (2020), 1–16.

[188] VLASKAMP, B. N., AND HOOGE, I. T. C. Crowding degrades saccadic search performance. Vision research 46, 3 (2006), 417–425.

[189] VON DER HEYDT, R., PETERHANS, E., AND BAUMGARTNER, G. Illusory contours and cortical neuron responses. Science 224, 4654 (1984), 1260–1262.

[190] WALD, G. Photo-labile pigments of the chicken retina. Nature 140, 3543 (1937), 545–546.

[191] WARREN, T. Microsoft rolls out colorful new windows 10 icons, 2020.

[192] WATSON, D. G., BLAGROVE, E., EVANS, C., AND MOORE, L. Negative triangles: simple geometric shapes convey emotional valence. Emotion 12, 1 (2012), 18.

[193] WEBB, J. M., SORENSON, P. F., AND LYONS, N. P. An empirical approach to the evaluation of icons. ACM SIGCHI Bulletin 21, 1 (1989), 87–90.

[194] WICKENS, C. D., GORDON, S. E., LIU, Y., ET AL. An introduction to human factors engineering. Longman New York, 1998.

[195] WICKENS, C. D., GORDON, S. E., LIU, Y., ET AL. An introduction to human factors engineering. Longman New York, 1998.

[196] WOBBROCK, J. O., FINDLATER, L., GERGLE, D., AND HIGGINS, J. J. The Aligned Rank Transform for non-parametric factorial analyses using only ANOVA procedures. In Conference on Human Factors in Computing Systems (New York, NY, USA, 2011), ACM, pp. 143–146.

[197] WOLFE, J. Visual search. pashler, h.(ed.), attention, 1998.

[198] WOLFE, J. M. Guided search 2.0 a revised model of visual search. Psychonomic bulletin & review 1, 2 (1994), 202–238.

[199] WOLFE, J. M. Visual attention. In Seeing. Elsevier, 2000, pp. 335–386.

[200] WOLFE, J. M. Visual search: How do we find what we are looking for? Annual Review of Vision Science 6 (2020).

[201] WOLFE, J. M., AND HOROWITZ, T. S. What attributes guide the deployment of visual attention and how do they do it? Nature reviews neuroscience 5, 6 (2004), 495–501.

[202] WOLFE, J. M., KLEMPEN, N., AND DAHLEN, K. Postattentive vision. Journal of Experimental Psychology: Human Perception and Performance 26, 2 (2000), 693.

[203] WU, L., LEI, T., LI, J., AND LI, B. Skeuomorphism and flat design: evaluating users' emotion experience in car navigation interface design. In Design, User Experience, and Usability: Design Discourse. Springer, 2015, pp. 567–575.

[204] XI, T., AND WU, X. The influence of different style of icons on users' visual search in touch screen interface. In International Conference on Applied Human Factors and Ergonomics (2017), Springer, pp. 222–232.

[205] YANTIS, S. Stimulus-driven attentional capture. Current Directions in Psychological Science 2, 5 (1993), 156–161.

[206] YOUNG, T. Ii. the bakerian lecture. on the theory of light and colours. Philosophical transactions of the Royal Society of London 92 (1802), 12–48.

[207] ZEKI, S. The functional organization of projections from striate to prestriate visual cortex in the rhesus monkey. In Cold Spring Harbor Symposia on Quantitative Biology (1976), vol. 40, Cold Spring Harbor Laboratory Press, pp. 591–600.

[208] ZHAI, S., AND KRISTENSSON, P.-O. Shorthand writing on stylus keyboard. In Proceedings of the SIGCHI conference on Human factors in computing systems (2003), pp. 97–104.

[209] ZHANG, X., WANG, Q., AND SHI, Y. Contrastive analysis on emotional cognition of skeuomorphic and flat icon. In China Academic Conference on Printing & Packaging and Media Technology (2016), Springer, pp. 225–232.

[210] ZHAOPING, L., AND LI, Z. Understanding vision: theory, models, and data. Oxford University Press, USA, 2014.

[211] ZHENG, J., BI, X., LI, K., LI, Y., AND ZHAI, S. M3 gesture menu: Design and experimental analyses of marking menus for touchscreen mobile interaction. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (2018), pp. 1–14.

[212] ZHOU, A. Cybernetics and human-computer interaction: Case studies of modern interface design. In 2014 IEEE Conference on Norbert Wiener in the 21st Century (21CW) (2014), IEEE, pp. 1–6.

[213] ZHOU, H., FRIEDMAN, H. S., AND VON DER HEYDT, R. Coding of border ownership in monkey visual cortex. Journal of Neuroscience 20, 17 (2000), 6594–6611.

APPENDIX A

CONSENT FORM

UNIVERSITY OF SASKATCHEWAN

Department of Computer Science
176 Thorvaldson Building
110 Science Place Saskatoon SK S7N 5C9 Canada
Telephone: (306) 966-4886 Facsimile: (306) 966-4884

Participant Consent Form

You are invited to participate in a research study entitled: *Effects of Visual Distinctiveness on Learning and Retrieval in Icon Toolbars.*

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, please ask. Please take the time to read this form carefully and to understand any accompanying information.

Researcher(s):

Dr. Carl Gutwin
Professor, Undergraduate Chair
Department of Computer Science
University of Saskatchewan
(306) 966 – 8646
gutwin@cs.usask.ca

Febi Chajadi
Graduate Student
Department of Computer Science
University of Saskatchewan
febi.chajadi@usask.ca

Purpose and Objective of the Research:

- This study is concerned with detecting user performance in tasks related to finding icons in a ribbon-like interface.
- The goal of the research is to examine the details of how icon design and affects the learnability and retrieval of icons.

Procedures:

- The session will require **60** minutes, during which you will be asked to complete a task involving finding several specific items within a small dataset utilizing a ribbon-like interface. You will then be asked to find those items again several times.
- You will be asked to fill out a demographic questionnaire before beginning the study, as well as asked to evaluate your performance after each round of the study, and finally you will be asked which approach you preferred.
- This study will take place in the Human-Computer Interaction Lab at the University of Saskatchewan.
- Please feel free to ask any questions regarding the procedures and goals of the study or your role.

Page 1 of 4

Department of Computer Science
176 Thorvaldson Building
110 Science Place Saskatoon SK S7N 5C9 Canada
Telephone: (306) 966-4886 Facsimile: (306) 966-4884

- At the end of the session, you will be given more information about the purpose and goals of the study, and there will be time for you to ask questions about the research.

Funded by:

- The University of Saskatchewan.
- The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks:

- Your hands/fingers/eyes may get tired using our computer software system.
- This will be addressed by allowing you to rest between every round as long as you like.

Compensation:

- As a way of thanking you for your participation and to help compensate you for your time and any travel costs you may have incurred, you will receive a **$10** honorarium at the end of the session.

Confidentiality:

- All personal and identifying data will be kept confidential. A participant ID number will be used, and no link will be kept between your participant ID and any identifying information about you.
- The anonymized data collected from this study will be used in articles for publication in journals and conference proceedings.
- **Storage of Data:**
  - This informed consent form and all research data will be kept in a secure location under confidentiality, in accordance with University policy, for 5 years after publication.

Right to Withdraw:

- Your participation is voluntary and you can answer only those questions that you are comfortable with.
- **You are free to withdraw from the study at any time without penalty and without losing any advertised benefits.**
- Withdrawal from the study will not affect your academic status or your access to services at the university.

UNIVERSITY OF SASKATCHEWAN

Department of Computer Science
176 Thorvaldson Building
110 Science Place Saskatoon SK S7N 5C9 Canada
Telephone: (306) 966-4886 Facsimile: (306) 966-4884

- If you withdraw, your data will be deleted from the study and destroyed.
- Your right to withdraw data from the study will apply until data has been pooled. After this, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up:

- As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of this study once they have been compiled (usually within two months). This summary will outline the research and discuss our findings and recommendations. This summary will be available on the HCI lab's website: http://www.hci.usask.ca/

Questions or Concerns:

- Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:
    - Dr. Carl Gutwin, Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

Consent:

Your signature below indicates that you have read and understand the description provided:

I have had an opportunity to ask questions and my/our questions have been answered. I consent to participate in the research project. A copy of this Consent Form has been given to me for my records.

Name of Participant	*Signature*	*Date*

______________________________ _______________________

UNIVERSITY OF SASKATCHEWAN

Department of Computer Science
176 Thorvaldson Building
110 Science Place Saskatoon SK S7N 5C9 Canada
Telephone: (306) 966-4886 Facsimile: (306) 966-4884

Researcher's Signature *Date*

A copy of this consent will be left with you, and a copy will be taken by the researcher.

APPENDIX B

DEMOGRAPHICS QUESTIONNAIRE

UNIVERSITY OF SASKATCHEWAN

Demographics

Demographic Questionnaire

Please answer the following questions. If you have any confusion in any question, please refer to the experimenter.

* **1. Participant ID:**

* **2. What is your age?**

* **3. Gender:**

- ◯ Male
- ◯ Female
- ◯ Non-binary
- ◯ Prefer not to disclose

* **4. Major or profession:**

* **5. Amount of computer use per week on average (hours):**

- ◯ 0
- ◯ 1-10
- ◯ 11-20
- ◯ 21-30
- ◯ 30+

* **6. What applications do you typically use in desktop/laptop?**

* **7. How do you issue frequently used commands?**

- ◯ keyboard shortcuts
- ◯ navigating through a GUI with the mouse
- ◯ Other (please specify)

*** 8. How often do you use keyboard shortcuts?**

- ◯ Always
- ◯ Often
- ◯ Sometimes
- ◯ Never

*** 9. How many keyboard shortcuts do you know?**

◯ 0	◯ 6-10	◯ 16-20
◯ 1-5	◯ 11-15	◯ 20+

APPENDIX C

STUDY 1 QUESTIONNAIRES

C.1 Concrete NASA-TLX

UNIVERSITY OF SASKATCHEWAN

NASA-TLX - Condition A

NASA - TLX

Please evaluate the task you just completed by carefully selecting the value on the scale from **0 (Low) to 7 (High)** at the point which matches your experience.
NOTE: *Performance* is measured on a scale where **0 is Poor and 7 is Good**.

* **1. Participant ID:**

* **2. Condition:**

- ◯ A
- ◯ B
- ◯ C
- ◯ D
- ◯ E

* **3. Mental Demand:**
How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) to perform the task?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

* **4. Physical Demand:**
How much physical activity was required (e.g., pressing, finger movement, controlling, activating, etc.) to perform the task? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

* **5. Temporal Demand:**
How much time pressure did you feel due to the rate at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 6. Performance:**
How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?

1 (Poor)	2	3	4	5	6	7 (Good)
○	○	○	○	○	○	○

*** 7. Effort:**
How hard did you have to work (mentally and physically) to accomplish your level of performance?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

*** 8. Frustration:**
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

9. Which targets were easiest to remember in this set?

- ☐ Border Vertical
- ☐ Text Rotate Up
- ☐ Copy to Clipboard
- ☐ Vertical Align Bottom
- ☐ Period Slider
- ☐ Rich Text Converter
- ☐ Requirements
- ☐ Not Interested
- ☐ Overscan Settings

10. Which targets were hardest to remember in this set?

- ☐ Border Vertical
- ☐ Text Rotate Up
- ☐ Copy to Clipboard
- ☐ Vertical Align Bottom
- ☐ Period Slider
- ☐ Rich Text Converter
- ☐ Requirements
- ☐ Not Interested
- ☐ Overscan Settings

C.2 Abstract NASA-TLX

UNIVERSITY OF SASKATCHEWAN

NASA-TLX - Condition B

NASA - TLX

Please evaluate the task you just completed by carefully selecting the value on the scale from **0 (Low) to 7 (High)** at the point which matches your experience.
NOTE: *Performance* is measured on a scale where **0 is Poor and 7 is Good**.

*** 1. Participant ID:**

*** 2. Condition:**

- ◯ A
- ◯ B
- ◯ C
- ◯ D
- ◯ E

*** 3. Mental Demand:**
How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) to perform the task?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 4. Physical Demand:**
How much physical activity was required (e.g., pressing, finger movement, controlling, activating, etc.) to perform the task? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 5. Temporal Demand:**
How much time pressure did you feel due to the rate at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 6. Performance:**

How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?

1 (Poor)	2	3	4	5	6	7 (Good)
○	○	○	○	○	○	○

*** 7. Effort:**

How hard did you have to work (mentally and physically) to accomplish your level of performance?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

*** 8. Frustration:**

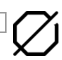
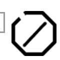
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

9. Which targets were easiest to remember in this set?

- ☐  Watch Later
- ☐  Unarchive
- ☐  Notes
- ☐  Closed Caption
- ☐  Caps Lock
- ☐  Volume
- ☐  Forum
- ☐  Contacts
- ☐  Documents

10. Which targets were hardest to remember in this set?

- ☐  Watch Later
- ☐  Unarchive
- ☐  Notes
- ☐  Closed Caption
- ☐  Caps Lock
- ☐  Volume
- ☐  Forum
- ☐  Contacts
- ☐  Documents

C.3 Concrete+Colour NASA-TLX

UNIVERSITY OF SASKATCHEWAN

NASA-TLX - Condition C

NASA - TLX

Please evaluate the task you just completed by carefully selecting the value on the scale from **0 (Low) to 7 (High)** at the point which matches your experience.
NOTE: *Performance* is measured on a scale where **0 is Poor and 7 is Good**.

* **1. Participant ID:**

* **2. Condition:**

- A
- B
- C
- D
- E

* **3. Mental Demand:**
How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) to perform the task?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

* **4. Physical Demand:**
How much physical activity was required (e.g., pressing, finger movement, controlling, activating, etc.) to perform the task? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

* **5. Temporal Demand:**
How much time pressure did you feel due to the rate at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

*** 6. Performance:**
How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?

1 (Poor)	2	3	4	5	6	7 (Good)
○	○	○	○	○	○	○

*** 7. Effort:**
How hard did you have to work (mentally and physically) to accomplish your level of performance?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

*** 8. Frustration:**
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

9. Which targets were easiest to remember in this set?

- ☐ Shuffle
- ☐ Pattern
- ☐ 360
- ☐ Find in Page
- ☐ Hearing
- ☐ Sleep Mode
- ☐ Ungroup Objects
- ☐ Uninstall Programs
- ☐ Remove Labels

10. Which targets were hardest to remember in this set?

- ☐ Shuffle
- ☐ Pattern
- ☐ 360
- ☐ Find in Page
- ☐ Hearing
- ☐ Sleep Mode
- ☐ Ungroup Objects
- ☐ Uninstall Programs
- ☐ Remove Labels

C.4 Abstract+Colour NASA-TLX

UNIVERSITY OF SASKATCHEWAN

NASA-TLX - Condition D

NASA - TLX

Please evaluate the task you just completed by carefully selecting the value on the scale from **0 (Low) to 7 (High)** at the point which matches your experience.
NOTE: *Performance* is measured on a scale where **0 is Poor and 7 is Good**.

*** 1. Participant ID:**

*** 2. Condition:**

- ◯ A
- ◯ B
- ◯ C
- ◯ D
- ◯ E

*** 3. Mental Demand:**
How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) to perform the task?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 4. Physical Demand:**
How much physical activity was required (e.g., pressing, finger movement, controlling, activating, etc.) to perform the task? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 5. Temporal Demand:**
How much time pressure did you feel due to the rate at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 6. Performance:**

How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?

1 (Poor)	2	3	4	5	6	7 (Good)
○	○	○	○	○	○	○

*** 7. Effort:**

How hard did you have to work (mentally and physically) to accomplish your level of performance?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

*** 8. Frustration:**

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

9. Which targets were easiest to remember in this set?

☐ Strikethrough	☐ Format Painter	☐ Password
☐ Select All	☐ Action	☐ Ink Editor
☐ Add Sticky Note	☐ Combine Files	☐ Watermark

10. Which targets were hardest to remember in this set?

☐ Strikethrough	☐ Format Painter	☐ Password
☐ Select All	☐ Action	☐ Ink Editor
☐ Add Sticky Note	☐ Combine Files	☐ Watermark

C.5 Mixed NASA-TLX

UNIVERSITY OF SASKATCHEWAN

NASA-TLX - Condition E

NASA - TLX

Please evaluate the task you just completed by carefully selecting the value on the scale from **0 (Low) to 7 (High)** at the point which matches your experience.
NOTE: *Performance* is measured on a scale where **0 is Poor and 7 is Good**.

*** 1. Participant ID:**

*** 2. Condition:**

- ◯ A
- ◯ B
- ◯ C
- ◯ D
- ◯ E

*** 3. Mental Demand:**
How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) to perform the task?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 4. Physical Demand:**
How much physical activity was required (e.g., pressing, finger movement, controlling, activating, etc.) to perform the task? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 5. Temporal Demand:**
How much time pressure did you feel due to the rate at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?

1 (Low)	2	3	4	5	6	7 (High)
◯	◯	◯	◯	◯	◯	◯

*** 6. Performance:**
How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?

1 (Poor)	2	3	4	5	6	7 (Good)
○	○	○	○	○	○	○

*** 7. Effort:**
How hard did you have to work (mentally and physically) to accomplish your level of performance?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

*** 8. Frustration:**
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

1 (Low)	2	3	4	5	6	7 (High)
○	○	○	○	○	○	○

9. Which targets were easiest to remember in this set?

- ☐ Diversity
- ☐ Move Grabber
- ☐ Used Items
- ☐ Reboot
- ☐ Deviations
- ☐ Extra Features
- ☐ Lock
- ☐ Omnichannel
- ☐ Answers

10. Which targets were hardest to remember in this set?

- ☐ Diversity
- ☐ Move Grabber
- ☐ Used Items
- ☐ Reboot
- ☐ Deviations
- ☐ Extra Features
- ☐ Lock
- ☐ Omnichannel
- ☐ Answers

C.6 Exit Questionnaire

UNIVERSITY OF SASKATCHEWAN

Exit Survey

1. Which set was easiest to remember?

2. Which set were you fastest with?

3. Which set did you make the fewest errors on?

4. Which icon set do you prefer?

5. What is your strategy for remember where the icons are?

APPENDIX D

PRESTUDY QUESTIONNAIRES

D.1 ColourFour NASA-TLX

D.2 ColourEight NASA-TLX

D.3 ColourTwelve NASA-TLX

D.4 GreyBox NASA-TLX

D.5 Exit Questionnaire

APPENDIX E

STUDY 2 QUESTIONNAIRES

E.1 UnfamiliarShape NASA-TLX

E.2 FamiliarShape NASA-TLX

E.3 Square NASA-TLX

E.4 Square+Colour NASA-TLX

E.5 Exit Questionnaire