Citation: Al Nuwaiser, W. (2020). Exploring the impact of icon similarity on user performance. (Unpublished Doctoral thesis, City, University of London)

Exploring the Impact of Icon Similarity on User Performance

Waleed Alnuwaiser

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

City, University of London
Centre for Human Computer Interaction Design
School of Mathematics, Computer Science & Engineering
Department of Computer Science

January 2020

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am most grateful to my first supervisor, Professor Stephanie Wilson, who provided me with considerable guidance and unlimited support. Professor Stephanie has always been there to provide feedback since she had started supervising me until the day of submitting this thesis. Professor Stephanie has always been supportive, helping me to overcome several obstacles that I had encountered during my research.

I am also very grateful to Dr. George Buchanan, who was my first supervisor before he moved to the University of Melbourne in Australia, from where he has continued to provide useful advice. I am thankful to Dr. George for being thoughtful, helpful, and for providing me with exceptional support throughout the years of my research.

I am thankful to my second supervisor, Dr. Ernesto Priego, who never hesitated to provide guidance and support for me whenever I needed it. Without the unstinting support of Professor Stephanie, Dr. Ernesto and Dr. George this thesis would have not been completed.

I would like to thank Stuart Scott, the interaction lab manager, for being helpful and cooperative in providing easy access to the interaction lab, and for lending me the eye-tracking equipment in order to complete this research.

I am indebted to my employer Al-Imam Mohammed Ibn Saud University in Riyadh, the Royal Embassy of Saudi Arabia in London and the Saudi Cultural Bureau in London for sponsoring my PhD programme at City, University of London.

I would like to extend my gratitude to all my colleagues in the Centre of HCID for their feedback, cooperation and sincere friendship, especially: Adrian Bussone, Ryanne Goodman, Tracey Booth, Reem Al Ashaikh, Carol Bulter, Timothy Neat, Aamna Toor, Axel Niklasson, Beatrice Vincenzi, Dara Sherwani, Alex Leigh, Anna Weller, Abi Roper, Gilang Pradana, Sondess Missaoui, Jordan Tewell, Graeme Aymer and Jonathan Day.

Finally, heartfelt thanks to my father, my mother, my brothers and my sisters for their love and enthusiastic support. Special thanks to my wife Basema Al-Saeed for her love, patience and for standing by me unconditionally always. Also, I am thankful to my son Mohammed and my daughter Hams for surrounding me with their kindness and joyfulness.

Declaration

I, Waleed Alnuwaiser, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been explicitly indicated in the text.

I grant powers of discretion to the City, University of London Librarian to allow the thesis to be copied in whole or in part without further reference to the author.

WALEED M. ALNUWAISER

Abstract

Modern mobile devices have relatively limited screen sizes and often use small icons in order to make effective use of screen space. Icons are visual elements that represent an underlying meaning or function. Icon designers must do their best to convey underlying meanings or functions using these small visual elements, but users may misinterpret the intended meaning of an icon. While previous research has studied how users interpret icons, the existing body of knowledge is surprisingly small. Previous research has investigated how users interpret icons based on characteristics such as visual complexity, concreteness, and semantic distance; however, researchers still do not fully understand what influences the success or failure of a user in interpreting icons or how the underlying interpretation process works.

The research presented in this thesis advances our understanding of how users interpret icons. It investigates how users make decisions regarding the meaning of an icon based on (1) seeing another icon that looks visually identical to it but has a different meaning (ambiguous icons) either before or after seeing the primary icon, (2) an icon's similarity or dissimilarity to a known 'target' icon, and (3) the visual context in which an icon appears. To address these questions, one exploratory study (Study I) and two experimental studies (Study II and Study III) were conducted. Studies I and II were lab-based, while Study III was conducted online. Study I found that users' ability to accurately interpret ambiguous icons is influenced by the order in which the icons are seen. These results are supported by qualitative data which show that users most frequently cite icon sequence as the reason for their interpretation of ambiguous icons. The second study showed a tendency for users' viewing times to increase as the visual dissimilarity of the icons increases. Furthermore, users tend to be highly accurate when interpreting the meaning of icons that look the same and have exactly the same meaning, while their accuracy is low when interpreting visually dissimilar icons. The findings of the third study suggest that the visual context has the potential to help users interpret icons. This thesis makes several contributions to the literature: It shows that the level of similarity or dissimilarity between icons influences users' performance and that the visual context may play a role in interpretation. Furthermore, it introduces a new method to determine the visual similarity or dissimilarity of icons. These findings have practical implications for the design of icons.

Chapter 1 Introduction

1.1 Background and Motivation

In the past 20 years, technology has become a part of everyday life. Current technology allows users to easily access many features and services through various mobile devices (e.g., laptops, phones, tablets, and wearable devices). As indicated by Statista (2018), mobile devices were used to view 49.7% of the total number of web pages accessed in the world in February 2017. Users of these devices want to be able to easily carry and use them while on the go. For example, users want smartphones that fit in their pockets or purses and laptops that fit in book bags. In order to provide users with devices which can easily be carried and used in various situations, companies have made the screen sizes of devices increasingly small (for example, a standard monitor is 1280 x 1024 pixels, an iPhone 3G is 320 x 480 pixels, and a handheld tablet is 800 x 600 pixels). However, this decrease in the screen size of devices has led to challenges in interaction, perception, and design.

Since mobile devices have smaller screens, designers must utilise this limited space efficiently and effectively. One way of achieving this is by creating icons. Unfortunately, the user's interpretation of an icon's meaning may not always be consistent with the designer's intended meaning (Derboven *et al.*, 2012). Users may misinterpret the designer's intended meaning of an icon for various reasons. One common reason for misinterpretation is that users have seen the icon before. For example, the heart icon on Facebook is used to 'love' a post. If a user has a Facebook page, he or she might assume that heart icons always mean 'love'. Thus, if the user visited a website for booking a hotel and saw a heart icon, he or she might assume that this icon could be used to 'love' the hotel. However, in a different context, a heart icon might be used for a different function. For example, on *Fab.com* a heart icon is used to 'like' a product and save/add it to a list of favourites. Another common reason for misinterpretation is that users are often influenced by personal factors (e.g., experience with internet browsing) or the meanings they associate with shapes or objects.

The following sections summarise prior research on the challenges that arise with the use of small screens, the importance of designing icons that users understand, and how users understand icons. This leads to the main research question of this thesis and the eight research questions that are addressed in order to answer the main inquiry.

The scope of the research questions is limited to interpretations of icons on small[1] screens, and this is reflected in the empirical work that was undertaken to answer the research questions. Although the results may have relevance to the interpretation of icons on other sizes of screens, the empirical studies did not investigate other screen sizes, and therefore, the findings cannot be assumed to hold true.

1.1.1 Challenges of Small Display Screens

Several studies have been conducted to investigate how small screens affect a user's ability to use devices. Jones *et al.* (1999) investigated web interaction problems related to small display screens. Two groups of participants were recruited to answer a questionnaire. One group was tested on a small screen, and the other group was tested on a large screen. The study found that the large screen group answered the questions correctly more often than the small screen group.

Duchnicky and Kolers (1983) conducted a study to explore the effects of display window height and text width on users' comprehension of reading passages. The authors investigated the readability of text on visual display terminals with three different line lengths, two different character densities, and five different window heights (either 1, 2, 3, 4, or 20 lines). They found that lines of full and two-thirds screen width were read, on average, 25% faster than lines of one-third screen width. Similarly, Kärkkäinen and Laarni (2002) found that displaying text on the small screens of personal digital assistants (PDAs), making the lines shorter, slowed down a user's speed of reading as manual scrolling and paging demand additional time.

More recently, Hancock *et al.* (2015) used a standard perceptual and cognitive test battery to examine the effects of the display size of smart watches on task performance. The authors conducted three experiments which examined the influence of varying viewing conditions on response speed, response accuracy, and subjective workload. In their experiment, four different screen sizes were used and there were three different levels of time pressure. The four screen sizes were 1. a PDA (320 x 280), 2. a handheld tablet (800 x 600), 3. a standard monitor (1280 x 1024), and 4. a large monitor (1600 x 1200). The results of this experiment were extensive, but the main point of concern for this review is that the authors found that small screen size had a negative influence on user performance when task demands were high.

---

[1] In this research, different viewport/display sizes were used: (1) iPhone 5s with 320 x 568 Pixels viewport size (4.00-inch screen size). (2) A mobile interface emulator with 375x667 Pixels viewport size which is equivalent to the viewport size of iPhone 7 (4.7-inch screen size).

Shrestha (2007) conducted a study comparing user effectiveness, efficiency, and satisfaction between a desktop browser and a mobile browser. Participants completed four tasks on both desktop and mobile browsers. Efficiency was measured by the total completion time for all eight tasks. Eighty percent of the total completion time was spent on the mobile browser, and 20% of the total completion time was spent on the desktop browser, indicating that tasks on the mobile browser were much more difficult than on the desktop browser.

A study by Raptis *et al.* (2013) compared three mobile phone screen sizes to investigate the effect of screen size on effectiveness, efficiency, and perceived usability based on the System Usability Scale (SUS). They found that screen size had a significant effect on efficiency. Users who interacted with screens larger than 4.3 inches were more efficient during information seeking tasks, indicating that their performance was better on larger screens.

Schade (2017) argued that the images displayed on large screens might not be directly suitable for display on small screens and discussed many problems that may arise when using large screen images on small screens. For instance, images with text may lead to unnecessary scrolling. Schade suggested that the text associated with images needs to be edited or removed to fit the limited space on small screens in order to minimise scrolling. For images to be moved from large to small screens, the size of the images or text has to be reduced. However, scaling images up or down can affect their clarity and readability. Determining the right sizes for images or text is important for the creation of clear and readable designs for mobile users.

As can be seen from these studies, small screen sizes can impair user performance. As suggested by Schade (2017), one way to deal with this issue is to be mindful of image sizes. This is an issue that designers are already aware of and have tried to combat using icons. Icons not only utilise the limited space of mobile devices effectively, but also allow quick access to complex meanings or functions (Gatsou *et al.*, 2011). However, while icons allow a user to operate more efficiently in the mobile environment, some issues still arise. When users interact with interfaces containing icons, they must determine the meaning of these icons. Hence, users are required to interpret and decode the meaning of icons to achieve their tasks through the interface.

1.1.2 Icons

Isherwood (2009) states that an icon is a small visual/graphical item that is used to convey a message to users through the interface of a computer system. An icon is a picture element of fixed shape with different size, colour, or texture attributes and may contain labels (Tepfenhart and Sowa, 1998). A user interface may contain selectable (clickable) or non-selectable (non-clickable) icons. A selectable icon can execute a function/program when clicked by a user, while a non-selectable icon can convey a meaning or message to a user, e.g., a logo. Both selectable and non-selectable icons were investigated in the research presented in this thesis.

1.1.3 Context

Context refers to any information that characterises the situation of an entity (Dey and Hakkila, 2008). Entities can be persons, places, or objects that affect the interaction between users and applications. Donald (2002) stresses that ensuring contextual relevance to users' tasks is contingent on achieving a high level of consistency and similarity between the users' requirements and the designers' intentions (Dey and Hakkila, 2008). Contextual perceptions, measures, and attributes are found to differ significantly between persons (Hiltunen *et al.*, 2005; Mäntyjärvi and Seppänen, 2002). Icons are always viewed within a certain context. Users may rely on different pieces of contextual information or resources to interpret the meaning of icons. In user interfaces, contextual information refers to the information that a user may obtain or rely on when navigating the interfaces and executing actions.

Contextual information may vary depending on how a user uses or navigates interfaces. For instance, information may be presented to users in a different order. In this research, context refers to the internal and external information or resources available to a user while navigating through web interfaces in a specific order. Internal context refers to the existing elements in a user interface and the website that is being studied (i.e., the surroundings of an icon being viewed or the undertaken task), whereas external context refers to elements outside the user interface and website (i.e., previous knowledge or other websites). In this research, the visual context of an icon is defined as the displayed visual content or information in a fixed image in which an icon appears.

1.1.4 Understanding the Meaning of Icons

The way in which a user interprets the meaning that a designer has assigned to an icon is important (Derboven *et al.*, 2012). Unfortunately, a user's interpretation of an icon's meaning may not always be consistent with the designer's intended meaning (Derboven *et al.*, 2012). Users may misinterpret an icon for various reasons. First, a user may interpret the icon as having the same meaning as an icon they have seen before. For example, if a user sees a thumbs-up icon, they may associate it with the ‘like’ button from Facebook. These types of misinterpretations arise because of the large number of icons that are used online and the various purposes in which they appear. Furthermore, a user’s interpretation may be influenced by many personal factors (e.g., experience with internet browsing) or the meanings they associate with shapes or objects. This leads to a subjective dimension of interpretation that complicates the recognition and understanding of icons. As such, individual differences in the interpretation of icons can lead to misunderstanding of icons’ functions (Dessus and Peraya, 2007).

For instance, various mobile operating systems may use icons differently (Facebook ‘like’ icon in Android will display a comment field, whereas in iOS will add the person who liked the post to the list of people who liked the post previously). One way of addressing this is to have consistency in graphical user interfaces, as this improves users’ performance (Reisner, 1981; Adamson, 1996). Reisner (1981) suggested a formal description as an analytical tool to discover inconsistencies in the design of interactive systems during the early stages before building a working/final model. Consistency of icons design has been introduced as a design principle/guideline in many software development platforms. For instance, Apple has developed design guidelines and principles for iOS developers to standardise user interfaces design and maintain consistent icons design. Thus, icons design must conform to a set of standards.

1.1.5 Recognising/Interpreting Icons

Previous research into the impact of icon characteristics has focused on three major characteristics: visual complexity, concreteness, and semantic distance (McDougall *et al.*, 2000; Shen *et al.*, 2018). Research showed that people respond more quickly and more accurately to simple icons than to complex icons (McDougall *et al.*, 2000), that people are more efficient at understanding concrete icons than abstract icons (Rogers and Oborne, 1987; Stammers and Hoffman, 1991), and that icons with less semantic distance (the amount of perceived distance between a picture and its meaning) are easier to identify than those with a large semantic distance (Goonetilleke *et al.*, 2001; McDougall *et al.*, 2001).

Furthermore, semantic distance has been shown to influence reaction times to icons; icons with larger semantic distances take longer to interpret (Blankenberger and Hahn, 1991; Cheng and Patterson, 2007). Cheng and Patterson (2007) had fourteen participants interact with icons which were grouped into three categories based on stereotypy. A response stereotype of an icon is defined as a most frequently occurring response to that icon. The degree of a stereotype is calculated as the ratio of the common responses to the total responses given by participants to an icon (Howell and Fuchs, 1968). These stereotypy categories were identifiable (60–100% identifiable), medium (30–60%), and vague (0–30%). The results showed that participants' reaction times increased as the identifiability of the icons decreased. Cheng and Patterson's results reflect the work of Rogers (1986) and Blankenberger and Hahn (1991) in that icons which are more concrete and have a small articulated distance require shorter reaction times and less cognitive processing than less concrete icons with a large articulated distance. The articulated distance of an icon is the relationship between the visual representation of the icon and its meaning. In other words, it means how well an icon is visually designed to represent its meaning.

Now we move from the level of icon characteristics to a broader level of users' understanding of icons. Research has shown that as users gain experience with icons, their performance at icon recognition and understanding improves (Green and Barnard, 1990; Stotts, 1998; Isherwood *et al.*, 2007). Hence, not only does the similarity between icons influence a user's interpretation of the icon's meaning, but the order in which icons are viewed may also influence this interpretation. If a user starts the experiment with an icon that they have not seen before they will likely be slow and inaccurate at interpreting its meaning; however, as they are presented with this icon throughout the experiment they will learn its meaning and become quicker at accurately recognising its meaning (Reber *et al.*, 1998; McDougall and Reppa, 2008). In fact, starting in the 1990s, researchers have shown that when presented with unfamiliar icons, participants perform poorly at recognising the icons' meanings (Haramundanis, 1996; Wiedenbeck, 1999; Isherwood *et al.*, 2007; Shen *et al.*, 2018). Exposure to the icons is necessary to induce retention of the concepts that icons represent (Wiedenbeck, 1999).

Another component which can influence how icons are interpreted is visual context. Haramundanis (1996) found that users interpret unfamiliar icons incorrectly and therefore additional information, such as supporting text, is required to enhance users' comprehension of unfamiliar icons. Huang and Bias (2012) showed that participants who are not familiar with icons and not given any context to understand the icons are inaccurate and inefficient at recognising icons. They also showed that participants required more time and made more mistakes when interpreting icons rather than textual information.

Huang *et al.* (2015) also showed with a functional magnetic resonance imaging study that while icons do stimulate the semantic system, they are not processed cognitively as logographical words. Instead, icons are processed more as images or pictures. As such, the authors concluded that icons are not as efficient as words in conveying meaning. Hence, it seems that users need additional information about the context in which an icon is being used in order to understand its meaning, especially if the icon is unfamiliar.

1.1.6 Conclusion

In summary, recent advances in technology have led to more and more people accessing the internet on devices of varying screen sizes. When screen size is limited, designers need to convey complex ideas in a small amount of space. One way that designers have combatted this issue is by using icons. The use of icons has helped to address the problems created by small screens; however, as users are confronted by more and more icons in daily life, it has become difficult for them to interpret the meaning of those icons as the designer originally intended (Silfver, 2012).

How users interpret icons is a complex subject which can be influenced by many factors, including characteristics of the icons and personal factors related to the user. When examining icons, research has focused on three major icon characteristics: visual complexity, concreteness, and semantic distance. The research in this thesis goes beyond these three characteristics by investigating visual similarity between icons and how it affects users' performance in understanding icons.

Additionally, the visual context in which icons are used and how it affects users' accuracy in interpreting the icons is investigated. The initial focus of this research is on the interpretation of ambiguous icons. ***Ambiguous icons*** are defined as icons that are visually identical but have different meanings. Next, the focus shifts to the *visual similarity* between a known 'target' icon and similar icons which must be interpreted, and the impact of the order in which users see the icons. This thesis introduces the use of *visual characteristics* to determine the similarity between icons, and then investigates how this similarity influences users' speed and accuracy in interpreting the meaning of icons. Furthermore, the effects of the presentation of visually similar icons are examined. Finally, the question of how icons are interpreted with and without visual context is investigated.

1.2 Research Objectives and Questions

The major goal of the research in this thesis is to investigate the factors at play in cases where new icons are encountered that may or may not bear a similar meaning to known ones. The central question that this research set out to address was: *How do users determine the meaning of visually similar and dissimilar icons?* This research examines a range of visual similarity levels between icons (icon type), from visually identical to entirely dissimilar. This enables a better understanding of how the visual similarity between icons affects users' understanding. The main research question is addressed by conducting one exploratory study and two experimental studies, which ask and answer eight research questions related to the main research question. Hence, in this research three research objectives are set out:

Objective 1: To explore the effect of presenting ambiguous icons in different orders on the accuracy of users' interpretations. This objective was addressed in Study I (RQ1.1 and RQ1.2). The analysis of Study I was inspired by the semiotic approach.

Objective 2: To identify the impact of icons' visual similarity on participants' speed of recognition and the accuracy of their interpretations (Study II; RQ 2.1, RQ2.2, RQ2.3, RQ2.4 and RQ2.5).

Objective 3: To explore the impact of the visual context in which an icon appears on the accuracy of users' interpretations (Study III, RQ3).

To meet these objectives, the following research questions for Study I, Study II, and Study III are identified:

- *RQ1.1: When viewing ambiguous icons, does the order in which icons are presented to participants influence how they interpret those icons? (Study I)*
- *RQ1.2: What reasons do participants give for their interpretations? (Study I)*
- *RQ2.1: Will participants who are presented with visually similar icons in a 'forward' order recognise the icons at a different speed than participants who are presented with visually similar icons in a 'backward' order? (Study II)*
- *RQ2.2: What is the impact of the degree of visual similarity between icons on the speed with which a participant recognises the icons? (Study II)*
- *RQ2.3: Will participants who are presented with visually similar icons in a 'forward' order interpret the icons with different accuracy than participants who are presented with visually similar icons in a 'backward' order? (Study II)*

- *RQ2.4: What is the impact of the degree of visual similarity between icons on the accuracy with which a participant interprets the icons? (Study II)*
- *RQ2.5: Is there a correlation between speed of recognition and accuracy of interpretation for each icon type (the four icon types)? (Study II)*
- *RQ3: What is the impact of visual context on participants' interpretations of the icons? (Study III)*

In Study I, visually identical icons with different meanings were tested to explore the impact of the order of icon presentation on the accuracy of users' interpretations. Study I investigated how users interpret/understand ambiguous icons that were viewed in two different orders. Study I was inspired by the semiotic analysis to characterise the interplay between user interface and user context. Insights were obtained into the impact of the specific sequential context in which icons appear on the subsequent interpretation of the icons, which aided in identifying the scale and nature of the effect.

In Study II, participants were presented with a known 'target' icon, then with a series of icons which varied in their similarity to the 'target' icon. The impact of visual similarity on viewing times and the accuracy of participants' interpretation of the icons was investigated. Furthermore, the order in which the series of icons was presented was studied. The order of presentation was chosen to begin with an icon identical to the 'target' icon and end with one that was completely dissimilar, or vice versa, giving two presentation order sequences. The speed of recognition and the accuracy of interpretation of the icons in these two sequences were compared to investigate the influences of similarity over time.

Study III expanded on Study II, as it was found that in Study II that participants were unexpectedly accurate on some of the visually similar icons. These results suggested that participants had become familiar with icons that were like ones they had already seen or known, and that this had an influence on icon recognition and understanding. Hence, the purpose of Study III was to test whether the visual context is important in the absence of familiarity. In Study III the participants were not presented with repetitive icons and, therefore, could not become familiar with the icons. Study III consisted of a 'visual context' group and a 'no visual context' group. The 'visual context' group were shown icons within their intended context by way of an image. The 'no visual context' group were shown the same icons; however, they were not given any contextual information.

1.3 Contributions

This thesis is primarily focused on how users understand/interpret the meanings of visually similar icons. This research contributes to the understanding of several aspects in the field.

1.3.1 The impact of different orders of presentation of visually similar icons on users' performance

The order in which icons are presented to users could play a role in the cognitive process by which users understand the meanings of the icons. Users usually navigate online websites or applications to achieve a goal. The navigation process typically involves a sequence of web pages, especially when the device being used has too little space to display the content all at once. Therefore, the impact of the order in which icons are presented to users on their performance was investigated in two studies.

In Study I, it was shown that presenting ambiguous icons in different orders can influence how users perceive them. The results indicated that the accuracy of users' interpretation of ambiguous icons varied (see RQ1.1 and RQ1.2 in Chapter 4).

In Study II, users were presented with visually similar icons in two different orders (order of icons) and did two tasks in two different orders (order of tasks). The results suggested that, when visually similar icons were presented in different orders, the presentation order had no 'main' effect on users' performance. (See RQ2.1 and RQ2.3 in Chapter 5).

As indicated by the findings of Study I, the accuracy of users' interpretation for ambiguous icons varied between the two orders. Therefore, a designer could avoid designing two icons that look the same but have different meanings, in order to reduce the confusion that affected users' performance. In other words, designing consistent icons is required to avoid users' confusion.

The results suggest that the order in which icons are presented to users is a key factor affecting how accurately the meaning of ambiguous icons is assessed. In turn, this naturally impacts the comprehension of these icons. Therefore, it might be useful to define a design technique to help designers increase the accuracy of users' interpretations of a sequence of icons presented within an interface.

1.3.2 How visual similarity of icons affects participants’ speed of recognition and the accuracy of their interpretations

The results of Study II showed that as the degree of visual dissimilarity between icons increases, participants’ speed of recognition decreases. This may indicate that participants need more attention and cognitive processes to determine the meaning of dissimilar icons (RQ2.2 in Chapter 5). Participants were more accurate at identifying identical icons than they were at identifying dissimilar icons (RQ2.4 in Chapter 5). Speed of recognition and accuracy of interpretation were not significantly correlated (RQ2.5 in Chapter 5).

The findings have relevance for design practice. As the visual dissimilarity of icons to a known ‘target’ icon increases, the speed with which users recognise icons decreases. Therefore, reducing the visual dissimilarity of icons to a known ‘target’ icon could be useful for designing icons in interfaces, making it easier for users to recognise icons as being like those they already know.

1.3.3 A method for determining the visual similarity between icons

This thesis introduces the concept of varying several visual fractions between two icons to determine the level of visual similarity between them (i.e., two icons are slightly different if there is one different visual fraction between them). Its contribution lies in the method for determining the degree of visual dissimilarity, from visually identical to entirely dissimilar, between an icon and a known ‘target’ icon. A visual fraction is defined as a single or compound shape that represents part of the whole icon’s shape (see Chapter 5).

1.4 Structure of the thesis

The structure of the thesis is as follows:

- Chapter 1 introduces the topic.
- Chapter 2 presents the methodology of the research conducted in the three studies.
- Chapter 3 presents a literature review of the topic and identifies the gaps in the research which this thesis fills.
- Chapter 4 presents the reasoning and findings of Study I, including the findings related to Research Questions 1.1 and 1.2.

- Chapter 5 presents the reasoning and findings of Study II, including the findings related to Research Questions 2.1, 2.2, 2.3, 2.4, and 2.5.
- Chapter 6 presents the reasoning and findings of Study III, including the findings related to Research Question 3.
- Chapter 7 presents an overall discussion and conclusion related to the three studies.

Chapter 2 Methodology

The purpose of this chapter is to introduce the research methodology used in this thesis. As stated in Chapter 1, the objectives set for the thesis were met and the research questions addressed through three empirical studies. This chapter discusses the concerns and decisions faced while planning each of these studies: the overall methodology that was used to conduct the research, the mixed-methods approach that was taken throughout this research, the research process, data analysis, and, finally, the ethical issues related to this research.

2.1 Methodology

The research discussed in this thesis was conducted using an empirical experimental methodology. One exploratory study and two experimental research studies were conducted in order to investigate the influence of the visual similarity of icons and icons' visual context on user performance. This approach allows for a deeper understanding of how the visual similarity of icons and icon context influence users' understanding of the meaning and function of icons. Furthermore, the research conducted utilised the mixed-methods approach, which is a combination of both qualitative and quantitative methods. From a philosophical standpoint, mixed methods are supported by the philosophy of critical realism (McEvoy and Richards, 2006). As a researcher, it is important to consider the philosophical perspective of one's research, because the perspective selected reveals the assumptions that the researcher is making about their research and influences the choices that they make about the purpose, design, methodology, methods, data analysis, and interpretation of their research.

Critical realism was developed by Roy Bhaskar in the 1970s (Braun and Clarke, 2013). Critical realism offers a solution to problems associated with combining quantitative and qualitative methods by providing a reconciliation at the ontological level (Scott, 2007). Critical realists distinguish between three different ontological domains or modes of reality (McEvoy and Richards, 2006; Bhaskar, 2013; Delorme, 1999). These modes are (1) the empirical, (2) the actual, and (3) the real 'deep' structures (McEvoy and Richards, 2006). The empirical mode refers to aspects of reality that can be experienced either directly or indirectly; the actual mode refers to aspects of reality that occur but may not necessarily be experienced; and real or 'deep' structures refer to the mechanisms that generate phenomena (McEvoy and Richards, 2006). Critical realists believe that the choice of methods should be dictated by the nature of the research problem and that the mixed-methods approach is the most effective approach for fully understanding these three modes of reality (McEvoy and Richards, 2006).

The mixed-methods approach is the most effective one for fully understanding the world around us (McEvoy and Richards, 2006; Olsen, 2002). Critical realists find the mixed-methods approach to be effective, because quantitative methods allow for the investigation of that which is predicted while qualitative methods allow for the investigation of that which is not predicted; combining these two methods allows the researcher to gather information about different facets of the same reality (McEvoy and Richards, 2006).

2.2 The Mixed-Methods Approach in User Research

The field of user experience has a wide range of research methods available, and these methods can be used to answer a wide range of questions. Rohrer (2014) discusses 20 methods which can be mapped across three dimensions: (1) attitudinal vs. behavioural; (2) qualitative vs. quantitative; and (3) the context of use. Figure 2-1 shows how the 20 user experience research methods map onto the three dimensions. It is not realistic to use all 2 methods in one study, nor does Rohrer propose this. User experience studies tend to benefit, however, when they combine multiple research methods, as this allows for combining insights. In order to determine which user research method, or which combination of methods, is appropriate, one must consider the questions that are being asked and the resources that are available to answer them.

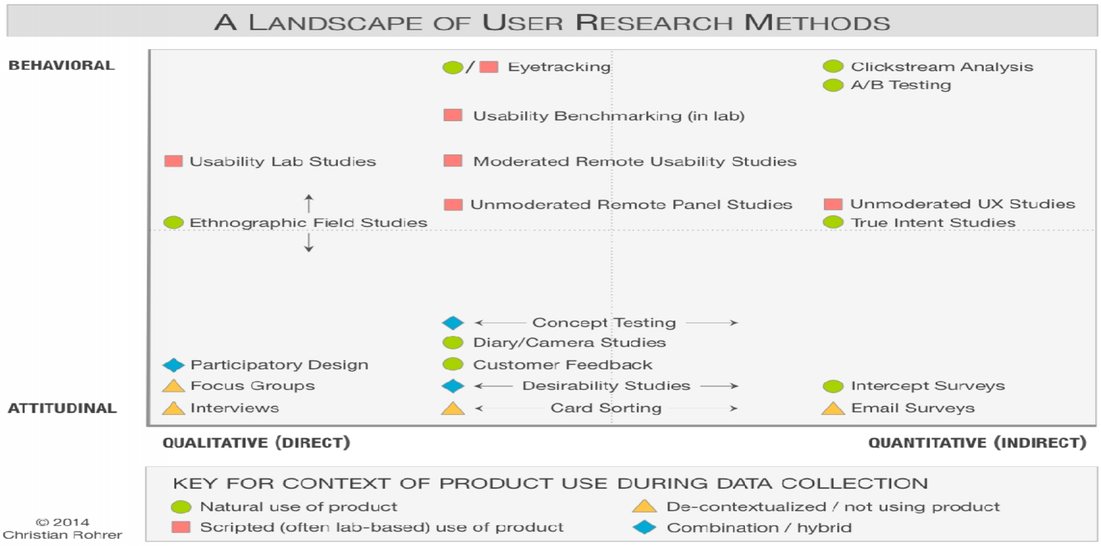


Figure 2-1: The 20 user experience research methods mapped across the three dimensions. (Image from Rohrer, 2014.)

According to Rohrer (2014), the distinction between the attitudinal and behavioural poles can be summarized by contrasting ‘what people say’ versus ‘what people do’. Attitudinal research is employed in order to understand or measure people’s stated beliefs, while behavioural research is employed in order to understand or measure ‘what people do’ with the product or service in question. Often data collected about what people say (attitudinal) and what people do (behavioural) can be quite different. Hence, combining these two methods can provide more information about what users do and why they do it.

Regarding the qualitative vs. quantitative dimension, according to Rohrer (2014) both research approaches are used to gather data about the behaviour or attitudes of users. Qualitative data, however, is gathered by observing the user *directly*. For example, in a field study a researcher directly observes how people use (or do not use) technology to meet their needs. Qualitative methods are used to answer questions about why or how to fix a problem. The qualitative methodology gives the researcher the ability to ask questions, probe people about their behaviour, or adjust the study protocol to better meet its objectives. Data collected during qualitative research is usually not analysed statistically. Quantitative data is gathered *indirectly* using measurements (e.g., eye-tracking data) or instruments (e.g., surveys with set answers). For example, a researcher may ask users to complete a questionnaire about their experience with technology and select their answers from a set (e.g., 1 = absolutely no experience 5 = a lot of experience). Quantitative methods are used to answer questions about ‘how many’ and ‘how much’. Quantitative methodology provides numbers which can help to prioritize resources and determine which issues have the biggest impact. Data collected during quantitative research is analysed statistically.

According to Rohrer (2014), the context of use has to do with how and whether participants in a study are using the product or service in question. The context of use can be described as natural, scripted, not using, or a hybrid of these three. When studying the natural use of a product, the goal is to minimize interference from the study in order to create an environment where the behaviour or attitudes of the users are as close to real-use scenarios as possible. Scripted studies are used when the researcher wishes to focus their research on specific usage aspects. Studies, where a product is not used, are conducted to examine issues that are broader than usage and usability, such as studies on a brand or larger cultural behaviours. Finally, hybrid methods can combine any of the methods.

2.3 Research Process

The studies conducted for this thesis utilised an empirical research approach which combined attitudinal and behavioural research, qualitative and quantitative data, and scripted contexts. The research combined different methods in order to collect more data about the nature of the users' experiences and behaviours. Each of the studies used some of the 20 user experience research methods discussed by Rohrer (2014): Study I used usability-lab studies and interviews; Study II used usability-lab studies, interviews, questionnaires, and eye-tracking; and Study III used remote usability studies. Rohrer (2014) defines usability-lab studies as ones which require the participant to come to the lab and interact one-on-one with a researcher on a given set of tasks which require the use of a product or service of interest. Rohrer (2014) defines interviews as one-on-one in-depth discussions between the user and researcher about the topic in question; eye-tracking as a method which uses an eye-tracking device to precisely measure where participants look as they perform tasks or interact naturally with websites, applications, physical products, or environments; and remote usability studies as those which are conducted online.

After a research problem for this thesis was developed, an overall question was formulated: *How do users determine the meaning of visually similar and dissimilar icons?* Based on this question, two initial research questions (RQ1.1 and RQ1.2) were formulated, and Study I was conducted to answer these questions. The results of the data analysis for Study I informed the hypotheses and research questions (RQ2.1, RQ2.2, RQ2.3, RQ2.4, and RQ2.5) that were posed in Study II. The information gathered during Study II and the ensuing data analysis and results informed the hypothesis and research question (RQ3) which were posed in Study III. The research process for this thesis, therefore, began with a question but the ways this question was investigated were informed by the data gathered during the experimental process. Figure 2-2 presents this chain of research processes. In the end, data from all three studies shed light on the original research problem and helped to provide a path for future research.

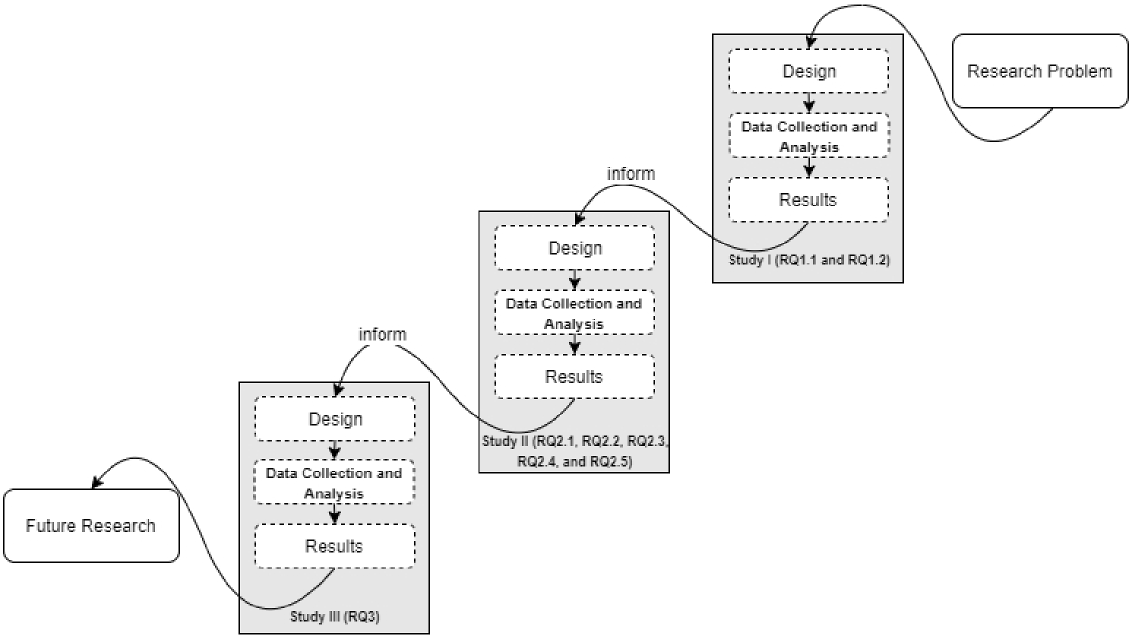


Figure 2-2: The research process chain.

2.4 Data Analysis

As mentioned in section 2.3, Studies I and II used interviews to collect qualitative data. The interview data were coded to aid the researcher in understanding the perspectives of the participants (Braun and Clarke, 2013) and also to produce quantitative data that the researcher could analyse using descriptive and inferential statistics. Both deductive and inductive coding were used.

2.4.1 Deductive Coding

All three studies required that the verbal/text data be deductively coded. Participants' interpretations of icons were deductively coded in order to analyse the accuracy of their interpretations. The coding terminology used for deductive coding was adopted from Islam and Bouwan (2016) who categorized participants' understanding of icons into five categories: accurate=5, moderate=4, conflicting=3, erroneous=2, and incapable=1. (For more details on the definitions of these codes, see section 4.2.4.1 in Chapter 4.)

2.4.2 Inductive Coding

Inductive coding was used in Studies I and II to further understand how the participants interpreted the meaning of the icons. During the inductive coding process, the reasons the participants reported for their interpretations of the icons were analysed.

2.5 Ethical Issues

Approval for all three studies discussed in this thesis was sought from the Computer Science Research Ethics Committee (CSREC) using the Ethics Proportionate Review Application. Once approval was given, the research studies were conducted. Standard ethics procedures were followed during each of the studies. Participants provided informed consent for all three studies. There were no foreseen risks to participating in this research and no adverse events were reported. All participants were over 18 years of age and did not demonstrate any impaired mental capacity. Meeting these inclusion criteria qualified them as participants in this study. All necessary precautions were taken to protect the participants' identities and information during the studies. Furthermore, to minimize any future risks related to confidentiality, all of the data collected during the three studies (including recorded audio materials, eye-tracking data, and text responses) will be retained for 10 years and then erased, as required by the University's Data Retention Policy.

2.6 Summary

This chapter discussed the methodology and the research methods used to answer the overarching research question and eight narrower questions. Chapter 3 presents a literature review of relevant prior work.

Chapter 3 Literature Review

3.1 Introduction

This chapter presents a review of the literature that influenced and informed the research conducted for this thesis. It informs the reader about the broad concept of this thesis, along with the issues the thesis intends to address. The recognition process of users, starting with seeing icons in a user interface and ending with them deciding on the meanings of these icons, is a complex one. The first section discusses insights, relevant concepts, and background for recognising and interpreting icons in user interfaces. The first section begins by discussing how icons could be ambiguous for users. Ambiguous icons are icons that different users may interpret differently even though they are presented with the same icon.

Next, previous work investigating the effect of limited-size screens on users' interpretation of icons is surveyed. Various factors that may influence users' performance, such as the visual similarity of icons and the context that icons appear in, are discussed. This section also addresses the role of Gestalt principles in working memory as well as the role of working memory in recognition and the recall of information. The second section focuses on various approaches that have been proposed for designing icons. Previous work investigating the characteristics of icons and how users identify such icons is presented. Next, the section discusses the semiotics approach and the ability of semiotics analysis to provide an effective method for recognising the relationship between a designer's intended meaning for an icon and a user's interpretation of the icon's meaning. The third section discusses the methods used to assess the visual and semantic similarity of icons. The final section focuses on eye-tracking techniques and the importance of the eye-tracker as a tool to record what users pay attention to (fixations). Finally, the cognitive processes which eye-tracking intends to access – attention and working memory – are discussed.

3.2 Interpreting and Recognising Icons

This section presents previous research on users' performance and the cognitive processes that underlie their ability to react to, interpret, and understand icons. The discussion starts by investigating how icons can be ambiguous for users, followed by looking at how users interpret icons that are viewed on limited-size screens. After that, the section discusses how users' performance might be influenced by the similarity or the dissimilarity between icons and the context of the icons. Cognitive insights are provided by discussing the role of working memory and Gestalt principles in users' understandings. Following subsections also review previous work that has investigated the speed of users' recognition. Finally, findings regarding the similarity between icons, the speed at which users recognise icons, the context in which icons are used, and the limited display of icons are synthesized.

3.2.1 Ambiguous Icons

In recent years, it has become commonplace for people to use various mobile technologies. With the development of these technologies, designers have had to create an increasing number of icons in order for users to effectively utilise the limited screen size of these devices. When a designer creates an icon, the goal is for the user to be able to interpret the meaning of the icon in the manner the designer intended. Not all users can transfer their knowledge from one device, app, or website to another, however: differences in interfaces and in icons between various devices, as well as differences between various apps or websites on the same device, can lead to divergent user interpretations. For example, imagine trying to switch from using a Windows laptop to an Apple laptop (or vice versa). If you have ever had this experience, you probably understand that it is not easy to switch between the devices of these two manufactures for many reasons – icons designs being one of them. Furthermore, some icons are ambiguous. They may serve multiple purposes and therefore can be interpreted in multiple ways.

As already stated above, the main goal of an icon is to help the user interact with a device without presenting the user with additional information or text. It is therefore important that the user can correctly interpret the meaning of an icon in order to use it effectively. How users interpret icons can be influenced by the devices they use, their own personal experience with icons, or the fact that the icon may represent multiple things. If icons are ambiguous, the user may not interpret them as the designer intended. One approach to understanding the difference between designers' and users' interpretations is to apply semiotic analysis. The semiotic analysis provides a means by which to understand how icons are designed and interpreted. As such, it allows the researcher to understand the communication process between the user and the designer, especially in the case of ambiguous icons.

3.2.2 Understanding Icons on Limited-Size Screens

In addition to icon similarity, the limited size of mobile device screens can influence how icons are interpreted. One reason that users may have trouble interpreting icons is the ever-decreasing size of the screens on which these icons are displayed. As users and life demand that technology, be more mobile and compact, designers have had to reduce screen size and display resolution (Burigat and Chittaro, 2011). This limitation arises since limited screen space reduces the number and size of symbols that can be displayed legibly on the screen (Stevens *et al.*, 2013).

The fact that small screens affect the user experience is supported by prior research, which tends to indicate that users can more easily navigate and use the internet on larger screens. Shrestha (2007) compared the ability of users to browse the internet on desktop and mobile devices. The study indicated that it took users nearly four times the amount of time to complete four tasks in a mobile browser than in a desktop browser. Other studies have shown that people answer questions more accurately on large screens than small ones (Jones *et al.*, 1999). These increases in time and reductions in accuracy may arise due to the small screen size making it hard for users to read the displayed text. Duchnicky and Kolers (1983) found that, on average, users read text 25% faster on larger screens than on smaller screens. These results are supported by Kärkkäinen and Laarni (2002), who found that the small screen size of Personal Digital Assistants (PDAs) slows down the speed of reading. If a user must utilise a mobile phone, research has shown that devices with screens larger than 4.3 inches allow the user to be more efficient during information-seeking tasks (Raptis *et al.*, 2013).

In recent years another type of device, which has an even smaller screen size than tablets, PDAs, or mobile phones, has entered the market: the smartwatch. While smartwatch screens tend to be smaller than the screens of the abovementioned devices, there are still differences in screen sizes among smartwatches. Research comparing smartwatches with small and large screens has shown that users find smartwatches with larger screen sizes to be easier and more enjoyable to use, resulting in these devices providing higher information quality than smartwatches with small screen sizes (Kim, 2017). These results are supported by Hancock *et al.* (2015), who compared user performance on smartwatches with performance on small and large screens and found that when task demands are high, user performance is negatively influenced by small screen size.

The effects of limited screen size on users' recognition and comprehension of icons have been investigated in the case of the official Android 4.0 and the official iOS

6.1 mobile operating systems (Lin and Lai, 2013). This work investigated the impact of visual form and the colour of icons on their recognisability and comprehensibility. Designing icons with a certain form and colour can quickly attract user attention and can assist them in comprehending the intended meaning (Lin and Lai, 2013). The primary focus of the study was on the visual representation of icons on mobile devices (which have limited-size screens) regardless of the icons’ context. The comprehension tests conducted provided significant evidence that many app icons should be improved or replaced in order to improve users’ interpretations of them.

Marcus (1996) has suggested that information can be easily recognised, understood, and remembered using metaphors, and therefore metaphors can be used to help enable users’ comprehension of icons on limited-size screens. The effective utilisation of metaphors is argued to allow users to understand, employ, and recall information more rapidly and easily, thus engendering increased user satisfaction through managing their expectations and comprehension (Roibás, 2002).

3.2.3 Visual Search on Limited-Size Screens

Mobile devices often have limited-size screens and contain many apps icons which look like each other. These factors can overwhelm the user as they search for the app icon, they are interested in using. Looking at this problem from a cognitive psychology perspective, users are essentially performing a visual search task every time they look for the app icon that they want to activate among many distracting apps icons. Visual search tasks are perceptual tasks that require attention. These tasks require a person to look at objects in their visual environment and find an object, target, among other objects, known as distractors (Treisman and Gelade, 1980). In a cognitive psychology laboratory, these tasks often consist of people looking for the letter L among numerous Ts; however, finding an app icon on a mobile device is a perfect real-world example of a visual search task. For example, imagine that you want to show someone a photo but there are three pages of apps icons on your phone, each app icon is the same size, and the photo album app icon is the same colour as five of your other app’s icons. All these factors will cause you to take longer to find the correct app icon.

Trapp and Wienrich (2018) were interested in exactly this issue and investigated how app icon similarity impacts visual search efficiency (how quickly a user can find an app icon) of users when they use mobile touch screen devices. The authors conducted two studies using knowledge from the field of cognitive psychology about visual search efficiency. More specifically, they considered three theoretical areas: (1) basic research regarding visual search, (2) applied research on the visual search of icons, and (3) the impact of (search) efficiency on user experience (UX).

One lesson from basic research regarding visual search that Trapp and Wienrich (2018) highlight is that of guidance by similarity. Prior research in cognitive psychology has shown that similarity strongly modulates the salience of a target, or, in plain English, how similar a target object is to the distractors around it modifies how easily a person can find that target. The more similar a target is to the distracting objects around it, the longer it will take a person to find it. On the other hand, the less similar a target is to the distracting objects around it, the more quickly a person will find it (Duncan and Humphreys, 1989). To return to the photo album app icon example, if the icon you are looking for is orange and five other apps around it are also orange, it will take you longer to decide which one is the photo album app icon. If the photo album app icon is blue and the five other apps icons around it are orange, it will not take you long to decide which one is the app icon you are looking for.

Within applied research dealing with visual search for icons, Trapp and Wienrich (2018) have discussed work by McDougald and Wogalter (2014) and Bzostek and Wogalter (1999). McDougald and Wogalter (2014) investigated the potential ability of colour to guide the user’s attention. They found that if relevant areas of pictograms are highlighted by colour, if there is no highlighting, users are able to provide more correct descriptions of the pictograms. From these findings, McDougald and Wogalter (2014) concluded that colour directs the user’s attention to the relevant areas of pictograms, which increases their comprehension of the pictograms. This work, however, only considered correct answers and not search time.

Bzostek and Wogalters (1999) study asked participants to rank coloured icons according to their noticeability. The authors used different icons and colours when presenting warnings at different screen locations. The authors found that when warnings were presented with coloured icons (blue and red), users were able to notice the warnings faster than when the warnings were presented with black icons. From this, Trapp and Wienrich (2018) concluded that colour can effectively guide attention to the relevant areas of the screen.

Based on these observations, Trapp and Wienrich (2018) conducted two studies to investigate visual search efficiency for app icon selection. In the first study, the authors varied set size (the number of app icons) and target presence as well as the visual similarity between icons using colour manipulation. The second study investigated visual search efficiency about the appeal of colourful icons and their effect on the perception of interaction qualities in terms of UX. While the findings from these studies are numerous, the ones relevant for this thesis can be highlighted. In the first study the authors found that when the target was easy to distinguish from the distractors, users were more efficient at finding the target than when the target was similar to the distractors. These findings are in line with prior cognitive psychology research. These results were also replicated in the second study, which had more participants ($N = 36$ versus $N = 18$) and was more realistic. The first study was more closely related to cognitive psychology experiments in the way the icons were presented, while in the second study the icons were presented in a way that is like how they would be displayed on a smartphone.

Another study investigating the influence of limited screen size on visual search tasks was conducted by Maurer *et al.* (2010). In their publication, Maurer *et al.* (2010) first discuss how mobile internet has become standard but how, due to bandwidth, input, and screen limitations, website providers often have to create special versions of their websites for mobile devices. The websites which are created for mobile devices are adapted by changing their size, design, and content to fit small screens and to require smaller amounts of data to be transmitted. Lam and Baudisch (2005), however, have found that reformatting websites to fit small screens distorts websites and often results in a lack of usability.

In order to investigate users' experiences of mobile websites on mobile devices, Maurer *et al.* (2010) conducted an online survey with 108 participants. The survey asked participants about their browsing habits and preferences on mobile devices. The first part of the survey was concerned with mobile device usage and the second part was concerned with mobile browsing on touch screen devices. Only users who had experience with touch screen devices completed the second part of the survey. The participants answered the survey questions on a Likert scale which ranged from 1 = never to 5 = very often.

The results of the survey indicate that one-third of the participants never use their phone for mobile internet access; however, 25% of the respondents used their phone for mobile internet access very often. A majority, 59%, had used a touch screen device before. The 63 respondents who had reported using a touch screen device before continued to the second part of the survey. These respondents reported having owned their mobile devices, on average, one year longer than those who said they do not have a touch screen device. More than half of the touch screen users had moderate to high levels of experience (4–5 on the Likert scale) with the iPhone or the iPod touch. Of the touch screen users, 57% owned a touch device themselves, with 19 % owning an iPhone.

In Maurer *et al.*'s survey, the participants were also asked to compare mobile-tailored and desktop-style websites. The participants were presented with two screenshots of the two different website versions and asked questions about the images. When given a choice between using the mobile-tailored version or the desktop one, 44% of the users preferred the mobile version, 30% had no preference, and about 25% preferred the desktop version of the website when using their mobile device. The participants who preferred the mobile-tailored website stated that it was easier to read and that they did not have to zoom in and out when using such pages. The participants who preferred the desktop version stated that it was easier to use because it looked similar to the browsing experience that they would have on a desktop computer. The participants also felt that the desktop version used on the mobile device allowed for sharing more information than the mobile version did. Both groups – those who preferred the mobile version and those who preferred the desktop one – felt that the version they preferred provided better clarity. Finally, an analysis of questions related to speed, simplicity of use, and clarity revealed that participants preferred the desktop version for speed but the mobile version for clarity and simplicity of use.

After the online survey was completed, Maurer *et al.* (2010) conducted a follow-up user study with 24 participants. In this study, participants performed a visual search task. More specifically, the users were asked to search for a keyword in a news article on a fictional news site using an iPod touch and the Safari web browser. Users interacted with both the mobile and the desktop version of the fictional news site. After the completion of these tasks, the users filled out a questionnaire similar to that used in the first study. The results of the survey show that 71% of the participants thought that they found the keyword faster when using the mobile webpage. Similarly, 66% of the participants felt that the mobile version was easier to use than the desktop version and 58% felt that the mobile version provided more clarity. Statistical analysis of the data revealed, however, that there were no statistical differences between the users' performance on the two website versions. While the results are not statistically significant, the study is still interesting because it shows that, when confronted with limited screen size, users have different preferences regarding how they interact with the websites that they view: some prefer mobile websites, while others prefer desktop versions. Furthermore, these preferences do not necessarily align with performance.

3.2.4 Limited Screen Size: Conclusions

As can be seen from the literature, limited screen size can cause issues for usability. This is since the limited screen size of mobile devices tends to lead developers to attempt to reduce the amount of information on the screen in order to keep the user interface simple, which results in the limitation of functionality. In order to create effective and efficient screen displays which do not overload the user with information, developers need to wisely create icons which clearly represent the necessary information.

3.2.5 Factors Affecting Users' Performance

This section discusses two factors that can affect users' performance in understanding the meaning of icons: (1) the similarity between the visual characteristics of icons, and (2) the context in which the icons are presented.

3.2.5.1 How the Visual Similarity of Icons Influences Users' Performance

The similarity or dissimilarity of icons might influence the speed at which icons are reacted to and cognitively processed. Research has shown that if users are given more time, their performance on icon recognition and understanding increases (Green and Barnard, 1990; Stotts, 1998; Isherwood *et al.*, 2007). This is likely due to the users seeing the icons repeatedly. When investigating icons, it is therefore important to consider their similarity.

The degree of similarity or dissimilarity of icons to a known 'target' icon might influence users' performance in recognising those icons. For example, in the beginning, a user may be slow to react to an icon and their accuracy in recognising the function of the icon may be low; however, over time and with repeated exposure to this icon, the user's ability to react to and identify the function of the icon will increase (Reber *et al.*, 1998; McDougall and Reppa, 2008). Research has shown that users tend not to be able to recognise the meaning of unfamiliar icons (Haramundanis, 1996; Wiedenbeck, 1999; Isherwood *et al.*, 2007; Shen *et al.*, 2018). In order for users to recognise the meaning of an unfamiliar icon, they must be exposed to it repeatedly to induce retention of the concepts that the icons represent (Wiedenbeck, 1999).

Singer and Lappin (1976) have demonstrated that the detectability of a difference between two forms depends on the context that the forms appear in. Lin and Luck (2009) have investigated the effect of similar representations on the performance of participants' memory. They found that memory performance improved for similar items through a colour-change detection task; compared to dissimilar items. Jiang *et al.* (2016) tested how the degree of similarity between three faces morphed together impacted memory performance. The three different morphs of similarity were randomly applied to faces using 30%, 50%, and 70% similarity. The study showed that high similarity facilitates memory performance by reducing the noise of representation in the memory. The similarity used in the study by Jiang *et al.* (2016), however, was designed using a single category, and the authors stress that similarity still could be a disadvantage for memory performance if the similarity between stimuli is highly different, such as comparing single-category similarity (i.e., only faces) with mixed-category similarity (i.e., faces and objects).

3.2.5.2 How Context Influences Icon Interpretation

Another factor beyond similarity, which can influence how icons are interpreted, is context. When icons are unfamiliar, users interpret them incorrectly; therefore, in order to enhance users' comprehension of unfamiliar icons, it may be beneficial for designers to add supporting text (Haramundanis, 1996). Huang and Bias (2012) have found that users respond inaccurately to and are inefficient at recognising icons which are unfamiliar. In fact, the interpretation of textual information by users requires less time and results in fewer mistakes than the interpretation of unfamiliar icons (Huang and Bias, 2012). These differences in processing textual information and unfamiliar icons may stem from the fact that icons are processed as images or pictures and are not processed cognitively as logographical words (Huang *et al.*, 2014). Huang *et al.* (2014) also suggest that although icons stimulate the semantic system, they are not as efficient as words in conveying meaning. Hence, it seems that users need more information about the context that an icon is being used in to understand its meaning, especially if the icon is unfamiliar.

It is important to consider how textual context can influence users' interpretations of icons, because in Study I of this thesis the users were shown icons along with their booking details in text (e.g., the length of the journey or the length of stay). In Studies II and III, however, the users were not only given textual information on web interfaces but were also presented with images (e.g., of a hotel lobby). It is therefore important to consider how images may influence users' interpretations of icons. The author found no prior studies that explicitly address how images may influence a user's interpretation of an icon. Work by Harrison (2003), however, suggests that when an image is displayed for advertising purposes, the image can be considered to be an analytical image, meaning that the image is 'asking' the person to view its attributes. Furthermore, Harrison (2003) suggests that images that are used for advertising can be symbolic; for example, an image of a motorcycle can represent virility.

Another important aspect of context is the placement of the icons in relationship to one another and images. People automatically encode spatial information (Mandler *et al.*, 1977; Hasher and Zacks, 1979) stated that this process tends to happen pre-attentively and may be used to reduce one's cognitive efforts in completing a task. For example, if a user is presented with an icon in the upper-right-hand corner, they may assume that the icon will be in the same location when they repeat the task. Prior research has shown that novice users of menu bars read the textual information presented in the menu; however, as users become more experienced with the menus, they no longer need to read the textual information but instead interact with the menu based on the spatial location of the function they wish to utilise (Kaptelinin, 1993). Furthermore, Moyes (1995) and Kaptelinin (1993) have shown that initially users focus on the local attributes of icons (e.g., form) but as users become more familiar with the webpage environment they begin to focus on the global attributes of the webpage/context (e.g., the icon's location compared to other icons) (Ark *et al.*, 1998).

3.2.6 How People Understand Icons: Deeper Insights into Working Memory and Gestalts Principles

In this section, the cognitive function of visual working memory (VWM) is introduced. Furthermore, Gestalt principles are introduced, as this set of principles suggests a framework for how visual items are processed. Considering the role of VMW and the Gestalt principles in visual perception provides an opportunity to gain deeper insights into how users understand icons. Finally, priming effects are discussed.

3.2.6.1 Working Memory

It is potentially useful to look at the recognition process users engage in when understanding the visual representations of icons and the role that working memory plays to support that understanding. Visual working memory provides short-term storage and allows the processing of relevant information from the visual environment, to account for temporary interruptions, such as saccades (Peterson and Berryhill, 2013). As such, cognitive processes are significantly anchored in VWM and it is important to investigate and assess the role of VWM and related cognitive processes in the interpretation and understanding of icons. One cognitive process that plays a role in interpreting and understanding icons is that of visual perception. Gestalt principles can be used to investigate and understand how visual perception works while a user interacts with an icon; they are especially useful for studying the principle of visual similarity, as it is one of the main aspects of this thesis.

3.2.6.2 Gestalt Principles

Gestalt principles were originally introduced by Wertheimer (1923), and since then many researchers have further developed them (Köhler, 1929; Koffka, 1935; Metzger, 1936, 2006). Gestalt principles, or Gestalt laws, describe how humans perceive scenes. Humans naturally perceive items in scenes, in organized patterns, and as objects. This innate disposition to perceive patterns in items seems to follow certain rules. These rules have been outlined in the Gestalt principles, which state that the brain places significance on organized structures rather than individual elements and places emphasis on emergent, holistic, and contextual perspectives (Soegaard, 2010). As such, when Gestalt principles are utilised, the brain may interpret unrelated objects as if they are one or belong together (Rock, 1986).

There are many Gestalt principles, but five of the main principles are discussed here: similarity, proximity, continuity, closure, and connectedness.

1. Similarity: Items that physically resemble each other are seen as part of the same object, and items that are physically different from each other are seen as part of a different object.
2. Proximity: Objects or shapes that are close to one another appear to form groups.
3. Continuity: The eye follows lines, curves, or a sequence of shapes in order to determine a relationship between these elements.
4. Closure: The brain automatically fills in gaps between elements to create a complete image.
5. Connectedness: Items that are connected to each other using colours, lines, frames, or other shapes are perceived as a single unit. Items which are not connected by any of these features are not perceived as a single unit.

How humans perceive groups of objects or parts of objects and form a whole out of these perceptions has constituted a significant strand of research within the field of visual perception (Soegaard, 2010). Early figure-ground phenomenon/Gestalt research revealed that some objects in the visual field assume a prominent position, while others are relegated into the background, resulting in visual segmentation (Rubin, 1915). This visual segmentation is characterized by the role of the human visual system in extracting and grouping similar elements and segmenting the scene into understandable patterns and positions of objects (Palmer, 1999). The visual system is known both for grouping items (Gestalt principles) and scenes (visual segmentation).

3.2.6.3 Prior Research on Visual Working Memory and Gestalt Principles

Understanding visual working memory and how Gestalt principles can play a role in it can provide us with information about how users interpret icons. VWM is essential for many cognitive processes; however, it is limited in capacity. This is where Gestalt principles come in. When a user interprets a visual scenario, their brain utilises Gestalt principles of grouping, especially the principles of connectedness, similarity, and proximity, in order to effectively use the limited resources of VWM. As the brain groups features together in order to interpret whole objects, it is allowing VWM to 'see' these objects instead of individual features. For example, if a user is presented with a colon and a left-open parenthesis, like this :), the user might interpret this as simply a face instead of a semi-colon and a left-open parenthesis. When the brain groups the two objects together into one, it needs fewer cognitive resources than it would to remember the two separate objects (Li et al, 2018).

Prior research has shown that that the Gestalt principles of connectedness, common region, and spatial proximity facilitate VWM performance in change detection tasks (Jiang *et al.*, 2000; Woodman *et al.*, 2003; Xu, 2002, 2006; Xu and Chun, 2007). In a change detection task, participants are presented with an array; then, after a short delay, they are presented with a second array. In the no-change condition, the second array is identical to the first. In the change condition, the second array differs by a single item. Participants must identify whether a change has occurred or not, and if they note a change, they are often asked to point out what changed. In a study by Woodman *et al.* (2003), when two stimuli were connected using the Gestalt principle of connection, accuracy improved by 6%. When the Gestalt principle of grouping by proximity was applied, accuracy improved by 12%. In a study by Xu and Chun (2007), stimuli which were grouped by common region resulted in higher VWM performance than ungrouped stimuli. Furthermore, research by Peterson and Berryhill (2013) has indicated that the principle of similarity can benefit VWM performance; however, the benefit of similarity on VWM performance was constrained by spatial proximity, such that similar items need to be near each other for participants' VWM to benefit from these groupings. In other words, not only does the Gestalt principle of similarity benefit visual perception, but it can also benefit VWM. This is important for the work discussed in this thesis as users may hold icons in their VWM, and in doing so they may unwittingly apply Gestalt principles to the icons in order to improve their ability to retain the icons. When conducting research on this topic, it is important to consider that the users' perception and interpretation of icons may be influenced by Gestalt principles and their VWM.

3.2.6.4 Attention and Gestalt Principles

While evidence shows that Gestalt principles can benefit VWM, research also suggests that the application of Gestalt grouping principles happens pre-attentively (meaning that it happens without consciousness) (Duncan, 1984; Duncan and Humphreys, 1989; Kahneman and Treisman, 1984; Moore and Egeth, 1997; Neisser, 1967; Mack and Rock, 1998; Mack *et al.*, 1992). When building patterns, such as ones created using Gestalt principles, bottom-up information-processing occurs. This means that features within the visual field are processed first; then patterns are found among these features; and, finally, these pieces are put together to create objects. These objects can then be retained in VWM (Duncan 1984; Neisser, 1967). Certain visual inputs are processed almost instantaneously and in parallel. This processing happens very early in the vision pathway and is pre-attentive. During pre-attentive processing the brain processes features such as orientation, length, closure, size, curvature, and colour, to name a few. This processing is done by the brain without conscious awareness, which results in the brain being able to use its cognitive abilities for other tasks. Put simply, pre-attentive processing allows the brain to have more resources to perform more complex tasks faster and more efficiently.

Research into how attention is used in the organization of visual information suggests that this parallel, pre-attentive processing serves to segment the visual field into separate objects. After the visual field is separated into objects, the process of focal attention can be applied; this process deals with only one object at a time (Duncan 1984, Neisser, 1967). These theories concerning visual information processing are supported by the work of Duncan (1984), who found that when participants must make two judgments about the same object, they can do so with accuracy; however, when participants must make two judgments about two different objects, they cannot do it accurately. This work supports both discrimination-based and space-based theories of visual attention. Discrimination-based theories propose a limit on the number of separate discriminations that can be made, and space-based theories propose a limit on the spatial area from which information can be gleaned.

Many researchers have shown that perceptual judgements about grouped elements can be made in the absence of attention (Driver *et al.*, 2001; Lamy *et al.*, 2006; Moore and Egeth, 1997; Russell and Driver, 2005). For example, Moore and Egeth (1997) had participants report which of two horizontal lines was longer; however, these lines were placed on different backgrounds. In experiments 1 and 2, the lines were presented with dots in the background. These dots were placed so that, if grouped, they could form displays like the Ponzo illusion which describes the mind's tendency to judge an object's size based on the background (Alleydog's Online Glossary, 2019). In experiment 3, the lines were also presented with dots in the background; however, these dots were placed so that, if grouped, they could form displays similar to the Müller-Lyer illusion which is an optical illusion that occurs when a line with two arrowheads pointing out is thought to be longer than a line with two arrowheads pointing in; however, the two lines are of equal length (Alleydog's Online Glossary, 2019). While the participants were inaccurate in reporting the patterns of the dots, meaning that they did not process the dots as patterns, their ability to perform the line-length discrimination task was affected by the two illusions. The author suggests that these results indicate that while Gestalt grouping can occur pre-attentively, attention is necessary for the groups created through Gestalt perception to be encoded in memory.

As can be seen from the literature, many studies support the theory that visual information is parsed according to the Gestalt principles of organization and that this process automatically enables visual perception (Peterson and Berryhill, 2013). In regard to the specific principles of Gestalt perception, research has shown that automated partitioning of images into meaningful regions and objects can also be applied using extraction, edge detection, and region extraction (Wang, 2003). Furthermore, there is strong empirical support that suggests that proximity has a greater facilitating effect for discrimination than similarity (Ben-Av and Sagi, 1995; Han *et al.*, 1999; Quinlan and Wilton, 1998). The inclusion of uniform connectedness in Han *et al.*'s (1999) study, however, suggested that the similarity principle provides benefits. Other studies have identified additive performance benefits by combining similarity and proximity (Kubovy and Van Den Berg, 2008). This strongly suggests that individual Gestalt principles are not equivalent, and it points to a gap in terms of a systematic hierarchy that evaluates the interactions of cue configurations.

3.2.6.5 Priming Effects

Major (2008, p. ii) defines priming as ‘the benefit that an event receives when its processing has been preceded by the processing of a related or identical event’. Bermeitinger (2014) has written a comprehensive review of priming in cognitive and social psychology. This section is based on Bermeitinger (2014) work and discusses priming in cognitive psychology. In cognitive psychology, priming paradigms are used to study memory or the pre-activation of concepts and motor reactions. Bermeitinger (2014) suggests a pyramidal structure for the levels of priming: at the bottom of the pyramid is the macro-level/perspective, in the middle is the midi-level, and at the top is the micro-level (although the borders between these levels/perspectives are fluid).

At the macro-level ‘each stimulus, each context, each action could be a prime that influences subsequent thoughts, actions, and feelings’ (Bermeitinger, 2014). It is typically assumed that macro-level primes pre-activate semantic concepts as well as longer-lasting motivational processes. At the midi level, primes may activate specific concepts, although these concepts are relatively global. Many memory- and recognition-priming paradigms work at the midi level. For example, when participants are given some words during an initial experimental phase and are then asked to produce words, they tend to produce words that are identical or semantically related to the words that were presented during the first phase. This effect occurs without instruction and at a higher rate than under control conditions.

The pre-activation of more specific concepts and actions occurs at the micro-level. At this level, researchers often present stimuli at very short time intervals, with the smallest time intervals being fractions of a second and the longest approximately two seconds. At this level, sequential priming is also often used. Sequential priming consists of a non-task-based prime that is presented for a maximum of a few hundred milliseconds and a target stimulus or target stimuli that are presented in rapid succession. In this type of paradigm, participants must often classify the targets based on certain criteria (e.g., as positive/negative, living/non-living, word/non-word, or left/right). Priming at the micro-level can be related to the pre-activation of specific concepts, reactions, goals, attitudes, or valences. This and the fast pace at which priming occurs at this level make micro-level paradigms good for investigating existing relations between different concepts, between concepts and actions, between concepts and attitudes, and so on.

At the macro level, one type of paradigm that researchers use is semantic priming. Typically, in semantic priming, the prime leads the participants to respond more quickly to semantically related concepts. Semantic priming paradigms can also be utilised at the micro-level. Many semantic priming paradigms use actual words; however, semantic priming can also involve a 'semantic relation based on shared attributes, such as the features shared by members of a given category (e.g., horses and cows are members of the category 'animals' and have in common that they are alive, have four legs, a tail, etc.)' (Bermeitinger, 2014).

Priming effects tend to be positive or negative. Positive priming effects occur when the participant responds faster to targets that are related to the prime; negative priming effects occur when the participant responds more slowly to targets related to the prime. Currently, positive semantic priming is believed to occur when features of the target overlap with the features of the prime. The overlapping features of the target are then pre-activated by the prime. In the case of negative semantic priming, some researchers believe that the target is used to clarify the identity of the prime. Other researchers believe that negative semantic priming is the result of an inhibition mechanism. According to this view, masked primes lead to a weak activation of representations, which causes the processing of these representations to be inhibited. When a target offers representations that are similar or the same as those that have been inhibited, the participant will take longer to respond to the target.

Priming effects have relevance to the research presented in this thesis when participants interpret the meaning of visually similar or identical icons. Priming effects might play a role in affecting participants' interpretations when they view an icon that is visually similar or identical to a recently viewed icon.

3.2.7 Recognition and Recall Processes

Recognition refers to the ability to recognise an event or a piece of information as being familiar, while recall refers to the retrieval of related details from memory (Budiu, 2014). In the research presented in this thesis, participants are asked questions related to their recognition of icons, which may require them to recall the details of icons. Understanding the processes of recognition and recall is therefore important. Prior research into recognition and recall has often used eye-tracking data to further understand these processes; this is also an important line of research as the present thesis utilises eye-tracking.

3.2.7.1 The Role of Memory in Recognition and Recall Processes

Previous research has shown that when participants are given more time to encode information into both their short-term (Pollack and Johnson, 1963; Norman, 1966) and long-term memory (Bulgelski, 1962; Murdock, 1960; Johnson, 1964), their performance on memory tasks is enhanced. Furthermore, studies have repeatedly demonstrated the reliability and vast storage capacity of long-term memory (Brady *et al.*, 2011; Bylinski *et al.*, 2015; Konkle *et al.*, 2010).

Research has also shown that images are encoded into memory differently depending on whether the expected upcoming task is a recall task or a recognition task. It seems that the brain uses different information to complete these two task types (Bahrick and Boucher, 1968; Frost, 1972; Tversky, 1973). It is important to note, however, that these memory-encoding benefits only occur when a person knows which task type, they will complete later. When the participant knows the future task type, their brain can prepare for the correct task type in advance, which results in improvements in performance (Tversky, 1973).

Work by Borkin *et al.* (2016) explicitly investigated how visualizations are recognised and recalled. They wanted to go beyond studying the memorability (how memorable something is) of visualizations and understand more about how visualizations are recognised and recalled. In order to do this, they labelled a dataset of 393 visualizations and analysed the eye movements of 33 participants on these visualizations. The authors found that titles and supporting text should convey the message of a visualization, that pictograms do not interfere with understanding and can improve the recognition of a visualization, that redundancy helps effectively communicate the message of a visualization, and that memorable content can allow visualizations to be encoded into memory at a glance. Overall, these results suggest that in order to understand a visualization, one must first pay attention to it and encode it in memory.

3.2.8 Speed of Recognition

When users try to recognise icons, their cognitive system employs recognition and recall processes. It is hard to practically and precisely distinguish between these processes; however, researchers have observed that the more familiar users are with icons, the faster their reaction times or speed of recognition is. These studies are discussed below. Section 3.3.5 presents some studies which address the speed of icon recognition. McDougall *et al.* (2000) have found that participants are quicker and more accurate in recognising simple icons than more complex ones. When comparing three categories of icons – identifiable (60–100% identifiable), medium (30–60%), and vague (0–30%) – Cheng and Patterson (2007) found that the speed of recognition increased as icon identifiability increased. Other researchers have shown that icons with larger semantic distance (the amount of perceived distance between a picture and its meaning) take longer and are harder to identify than icons with less semantic distance (Goonetilleke *et al.*, 2001; McDougall *et al.*, 2001; Blankenberger and Hahn, 1991). Researchers have also shown that less concrete icons with larger articulated distance require more time and cognitive processing to recognise than icons which are more concrete and have a small articulated distance (Blankenberger and Hahn, 1991; Rogers, 1986).

3.2.9 Tying Similarity, Speed of Recognition, Context, and Limited Screen Size Together

Gatsou *et al.* (2012) have discussed the impact of mobile interface icons on user interaction. The authors begin by discussing the necessity of icons and graphic elements for the use of mobile devices. Icons and graphic elements provide a means of effective communication between the device and the user via the user interface. It is imperative to have well-designed icons and graphics as they allow the user to recognise the function of the icons and graphics without additional instructions. It is important that designers create icons who's meaning the end-user can easily interpret and understand. As such, the design of icons becomes a semiotics task as designers use icons to represent specific meanings and the user is required to interpret this meaning. Furthermore, the link between the designed icon and its intended meaning should be obvious to all the users of the icon, and all these users should only interpret the icon in one way. Taking these things into consideration, Gatsou *et al.* ask the following questions:

1. Are mobile phone function icons easily recognisable by a wider audience?
2. Is there any difference in the recognition rate between different age groups?
3. Are there any differences in the recognition rate between the genders within each of the age groups?

In order to answer these questions, the authors recruited 60 participants between the ages of 20 and 79, of whom 32 were male and 28 were female. The participants completed a paper-and-pencil icon recognition questionnaire which consisted of 54 mobile phone icons. The results of this questionnaire revealed that for 29 icons the recognition rate was over 66.7%. According to the International Organization for Standardization (ISO3864), icon recognition rates should be at least 66.7% to be acceptable. Based on this, the authors graded 29 of the icons as ‘good’ and 25 of the icons as ‘bad’ (correct response rate below 66.7%) based on the ISO standard (Gatsou *et al.*, 2012). Further analysis also revealed that six of the icons were easily recognised and associated with their correct functions; however, these recognition capabilities were seen among the younger participants, while the older participants were less accurate in recognising and interpreting the meaning of the icons. Finally, the authors did not find a significant difference between the icon recognition rate of men and women.

Based on the authors’ findings (Gatsou *et al.*, 2012), they make the following six observations which should be considered when designing icons: ‘(1) Combinations of graphics that are complex or ambiguous decrease the ease with which the icon is correctly interpreted. (2) The use of familiar metaphors increases the likelihood that an icon will be interpreted correctly. (3) Users have trouble in correctly interpreting icons that employ symbolic or abstract representations. (4) Icons that employ concrete imagery are more frequently interpreted correctly. (5) Users draw upon their experience of the real world to interpret the functions conveyed by icons. (6) The scale of the screen size on which the icon is displayed influences how far the user correctly interprets the icon’ (Gatsou *et al.*, 2012).

Each of these findings and observations might be important for this thesis; however, some points are important. Observations 1, 2, 5, and 6 are important, as this thesis addresses the effect of visually similar icons, icons’ presentation order, and limited screen size on the ability of users to recognise and interpret icons.

3.3 Icons: Design and Characteristics

3.3.1 Designing Icons

This section discusses several ways in which icons can be designed: presentation discovery, semantic labelling, and the learning-based similarity metric of design. Section 3.3.2 informs the reader of the role of an expert in icon design. This is important because this thesis also employed experts to assess the design of similarity between icons. Section 3.3.3 discusses the importance of semantic labels in determining the similarity of icons, an important concept as this thesis investigates the effects of the visual similarity of icons on users’ perception and understanding of their meaning. Furthermore, users’ understanding of icons is analysed using semantic labels. Finally, section 3.3.4 discusses learning-based similarity metrics which show how icon sets are created based on similar visual style and identity.

3.3.2 Designing Icons Using Presentation Discovery

Payne and Starren have discussed presentation discovery, which is a structured approach to designing icons: ‘Presentation Discovery consists of four major steps: (1) The identification of target domain concepts for use in a presentation model, (2) The elicitation of candidate graphical primitives that represent the selected domain concepts from domain experts, (3) The categorical sorting of candidate graphical primitives into consensus clusters based on their visual characteristics, and (4) The extrapolation of representative prototype graphics from the consensus clusters’ (Payne and Starren, 2005, pp. 338-339).

After discussing the presentation discovery method, the authors state that their study focuses on steps 3 and 4 of the methodology (Payne and Starren, 2005). In their study, six domain experts were given 50 common textual mammography findings and asked to draw a graphic representation of those findings. After the experts created these graphics, non-domain experts sorted the resulting graphics into groups based on their visual characteristics. Sorter agreement was measured at both the individual graphic and concept-group level using a novel simulation-based method. Additionally, consensus clusters of graphics were derived using a hierarchical clustering algorithm. The results of this study show that non-experts can reliably group graphics based on the similarity of their underlying domain concepts. The process the authors used resulted in consensus clusters which provided graphic primitives informative for the design of icons.

3.3.3 Designing Icons Using Semantic Labels

Lagunas *et al.* (2018) have discussed procedures for designing an effective icon set. Designers (or researchers) tend to first provide semantic labels for icons. This ensures that the icons fit their intended meaning. Researchers have generated icons sets based on semantic labels and semantic relationships (Setlur *et al.*, 2005; Setlur and Mackinlay, 2014). More specifically, Setlur and Mackinlay (2014) developed a method for mapping categorical data to icons, while Lewis *et al.* (2004) have used shape grammar, an icon generation algorithm, to generate visually distinctive icons.

After providing semantic labels for icons, designers check that icons are visually appealing regarding style and visual identity. Style can be defined as the set of pictorial features of icons such as strokes, fill, or curvature, while visual identity refers to the higher-level properties which make a set of icons visually identifiable and unique. These higher-level properties are usually linked to the shape of the icons. Hence, icons can have a different visual identity while their pictorial style can be considered similar.

3.3.4 Designing Icons Using a Learning-Based Similarity Metric

Lagunas *et al.* (2018) have proposed a learning-based similarity metric which uses the properties of style and visual identity to help designers in the process of icon set creation. The authors greedily gathered an icon dataset from the Noun Project[2] online database. This database contains icons which are labelled with keywords and therefore have a semantics attachment. Furthermore, the icons are organized into collections which share a style and have a visual identity. Figure 3-1 shows an example collection that was gathered using this method. Style is defined as the set of the icon's visual features, and visual identity as the property that makes a collection of icons visually identifiable, such as the shape of the object (Lagunas *et al.*, 2018).

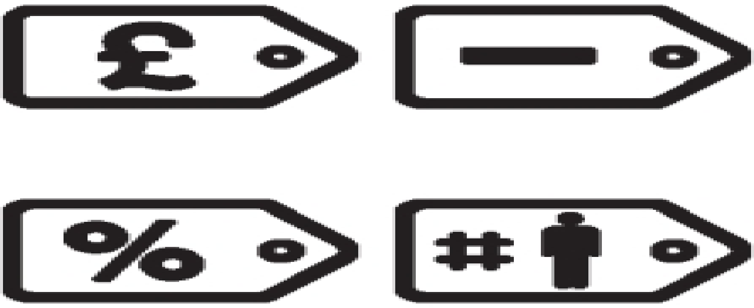


Figure 3-1: An example 'label' collection from the data set that was greedily gathered based on a style and visual identity. (Image from (Lagunas *et al.*, 2018).)

The authors then trained a Siamese neural network with an online dataset of icons which were organized in visually coherent collections which were used to adaptively sample training data and optimize the training process. The researchers also collected perceptual ratings from humans about the icons' similarity. These ratings were used to evaluate and test the authors' proposed model. The authors found that their method was able to model the high-level properties of the icons, capturing the icons' visual identity.

---

[2] https://thenounproject.com/

3.3.5 Three Icon Characteristics

After the above brief survey of the methods used to design icons, this section presents previous work and studies that have investigated the impact of icons' characteristics on how they are identified and understood. Researchers have tended to focus on three characteristics when discussing the visual characteristics of icons: visual complexity, concreteness, and semantic distance (McDougall *et al.*, 2000). McDougall *et al.* (2000) have shown that simple icons, in comparison to complex icons, result in users having quicker reaction times and a higher accuracy of icon interpretation. Furthermore, Cheng and Patterson (2007) found that reaction times increased as the identifiability of the icons decreased when three categories of icons were compared. The compared three categories were: – identifiable (60–100% identifiable), medium (30–60%), and vague (0–30%). Other researchers have shown that users are more efficient at understanding concrete icons than abstract ones (Rogers and Oborne, 1987; Stammers and Hoffman, 1991).

Prior research has also shown that icons with larger semantic distance (the perceived distance between a picture and its meaning) are harder to identify than icons with less semantic distance (Goonetilleke *et al.*, 2001; McDougall *et al.*, 2001). It has also been shown that icons with smaller semantic distance require less time to react to and that users are faster in interpreting them than icons with larger semantic distance (Blankenberger and Hahn, 1991 via Cheng and Patterson, 2007). Blankenberger and Hahn (1991) have also shown that less concrete icons with larger articulated distance require longer reaction times and more cognitive processing than icons which are more concrete and have a small articulated distance; these findings replicate work done by Rogers (1986).

3.3.6 Icons in Semiotics

This section discusses what semiotics is and how it can be used to understand a user's interpretation of the meaning of icons. Semiotics is often called 'the study (or theory) of signs' (Pu, n.d., p. 2), and it can be defined as the relationship between a sign, an object, and a meaning (Littlejohn, 1999, p. 62). Semiotics focuses on how signs are created, and the ways audiences understand those signs (Littlejohn, 1999, p. 330). As will be discussed further in the following paragraphs, icons are a type of sign that designers use when presenting information on the internet.

Icons are used by designers to represent complex concepts in a limited amount of space. Semiotic theory relates to icons because icons are a type of sign which is designed to represent a deeper meaning, and furthermore because it is important that users who interact with these icons understand the designer's intended meaning.

De Saussure's model of the sign posits a dyadic relationship: the sign can be represented by a signifier (i.e., an image) – which is the form that a sign takes – and a signified – which is the concept that the signifier represents (De Saussure, 1916).

Peirce classified signs into three different categories: icon, index, and symbol. The simplest of these categories is the icon, which is a visual representation of what it stands for. An index indicates or correlates to what is being represented (e.g., using a visual representation of a dark cloud to indicate the meaning of rain). A symbol embodies a conventional and cultural relation between the meaning and the sign. Since symbols have no logical or direct relation to the intended meaning, previous experience or cultural knowledge is required for a correct interpretation (Peirce *et al.*, 1931).

Figure 3-2 illustrates the relations between the three parts of the semiotic triangle that is discussed in Zakia and Nadin (1987): R for Representamen (the user interface), O for Object (program or function), and I for Interpretant (the user's interpretation).

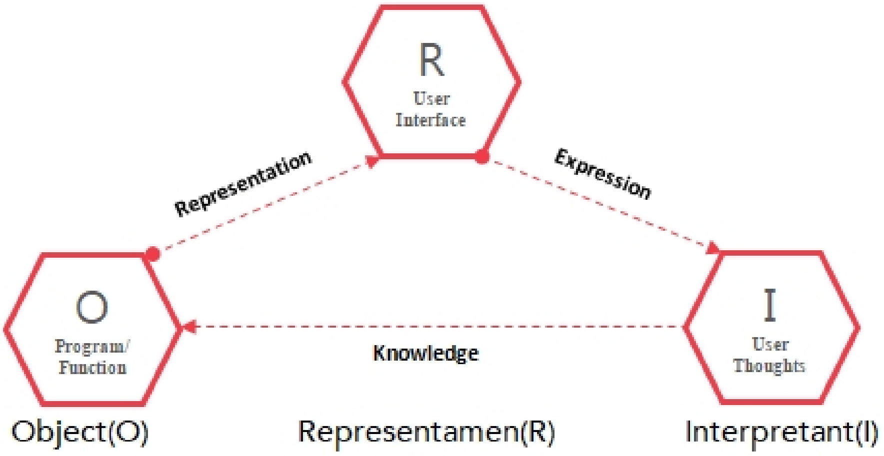


Figure 3-2: Semiotic triangle. (Image from Zakia and Nadin (1987).)

Icons display features that resemble the object they signify (Gatsou *et al.*, 2012). In a user interface (Ali *et al.*, 2015), icons are an important element of navigation which visually represent an object, idea, or action (i.e., a hamburger icon to represent a list or menu) (Bedford, 2014). A seminal triadic definition of the icon by Pierce (1931) provides insight into the issues underpinning its application. Firstly, the icon can be perceived in relation to its internal qualities that project a resemblance or analogy (e.g., a picture of a computer or person); secondly, it can refer to an entity through its external association or purpose (i.e., a flame denoting a fire hazard); and thirdly, it can be perceived in relation to how it is interpreted (Gatsou *et al.*, 2012). The second and third components of this definition implicitly support the importance of the semiotic approach in connection with the external world and in placing the icon in its external, relational context (Landriscina, 2013).

According to the Peircean model of signs, semiotics involves a dynamic process that is both context-sensitive and interpreter-dependent (Queiroz and Merrell, 2006). For instance, the meaning of two intersecting lines, ‘X’, can vary depending on the context. Unaccompanied or embedded in text, it can be perceived as the letter ‘X’; however, when presented with an error message and/or in the colour red, it can be perceived as an indication of malfunction or restriction. Furthermore, when it appears as a button, it can be perceived as an indicator of where to click in order to close, delete, or remove items (Stickel *et al.*, 2014). As can be seen from this example, in some instances even two visually similar icons might be interpreted in different ways. It is therefore important to understand how visual similarity, or dissimilarity, can influence a user’s perception and interpretation of an icon’s meaning.

3.3.7 Semiotic Frameworks in Human-Computer Interactions (HCI)

Semiotics has been brought to the attention of scholars studying human-computer interactions through many studies. For instance, Semiotics Interface sign Design and Evaluation (SIDE) provides a semiotics framework in order to maximize the usability of user interfaces (UIs) (Islam and Bouwman, 2015). Icons displayed on user interfaces have been studied and researched using empirical data, and they have been modelled at various levels of semiotics theory: syntactic, pragmatic, social, environment, and semantic. Especially the accuracy of users’ interpretations and icons’ level of intuitiveness have been studied and analysed for websites in a desktop environment (Islam and Bouwman, 2015). None of the context-specific impacts on the accuracy of users’ interpretation, however, have been studied or investigated using the SIDE framework.

The semiotic inspection method by de Souza *et al.* (2006) encapsulates interactive principles during five key stages to evaluate the ‘communicability of interactive computer-based artifacts’ (de S Reis and Prates, 2012): (1) the inspection of documentation and help information; (2) static and (3) dynamic interface signs; (4) a comparison of designer-to-user communications; and (5) conclusive appreciation of the overall quality of designer-to-user metacommunication (de Souza *et al.*, 2006). In effect, de Souza *et al.* (2006) have deconstructed the metacommunication process and the messages that describe this interplay. This deconstruction implies an understanding of who the users are and of their needs, preferences, and motivations, supported by the qualitative and interpretative processes posited by semiotic engineering methods.

Web-semiotic interface design and evaluation (W-SIDE) (Speroni, 2006) is a framework proposed to model and evaluate signs (semiotic units) in information-intensive web interfaces. Despite its name, W-SIDE is only remotely connected to SIDE and predates it. W-SIDE was proposed as a solution to bridge the gap between user confusion, caused by the pre-supposed knowledge of users when web signs are designed, and the real knowledge possessed by users. Speroni has argued that the sign in a web interface context conveys two layers of meaning: a content meaning layer, which refers to the prior knowledge of users about the domain of the web user interface, and a functional meaning layer, which refers to the action part of the web interface when the interaction takes place.

W-SIDE introduces a set of concepts, called an ontology, to improve users' knowledge to help them interpret the meaning of signs in web user interfaces (UIs). One of the ontologies proposed in the W-SIDE framework is the context ontology, which refers to providing knowledge of the specific context a sign appears in. It is limited to information that is not explicitly relevant to the website domain but that has an implicit relation to making dialogue possible and comprehensible. For example, in W-SIDE, the authors have included a teachers' section on the Getty Museum's website to help teachers find an educational resource even though the topic the website discusses is art. The W-SIDE framework, however, does not consider the context of interaction.

Others have used semiotics in more novel areas of HCI. Derboven *et al.* (2012) investigated users' understanding of a multi-touch interface on a table top application platform (MuTable) from the perspective of semiotics. In-depth analysis of the MuTable interface was conducted using De Souza's (De Souza and Leitao, 2009) communicability evaluation method (CEM). This analysis moved from low-level observations to high-level semiotic profiling in three stages: (1) tagging the problems that users encounter through a predefined coding scheme, (2) completing the interpretative stage, which refers to looking for problems that appear through metacommunication between the user and the designer; and (3) conducting semiotics profiling, which refers to the evaluation of how well the designer's message is being transmitted to the user (Derboven *et al.*, 2012). When users interact with the MuTable interface, their reactions are categorized using communications tags proposed in the study. For instance, 'Why doesn't it?' is a tag categorized as 'users seek to clarify the designer deputy's intended signification'. Users ask this question they wonder about why this part of the interface is not reacting as expected (Derboven *et al.*, 2012). Since the occurrence of the ‘Why doesn’t it?’ tag was extremely frequent compared to the other two tags, ‘Oops!’ and ‘What’s this?’, three different context-specific situations of ‘Why it doesn’t it?’ tag were analysed. These three situations were divided into occurring in the gesture context (a gesture problem), navigation context (a problem occurring when a user tries to open or navigate between widgets) and meaning assignment context (a situation where the outcome is different than what the user expected).

3.3.8 Assessing the Design of Icons

Now that the design of icons has been discussed, it is important to address how users’ understanding of icons can be assessed. ISO 9186 (Public Information Signs) provides a comprehensive methodology for ensuring that graphic symbols and signs can be easily recognised and understood by a general audience (ISO9186-1:2014, 2015). In the comprehension test laid out in the ISO standard, neither supplementary information nor contextual information are shown with a symbol. Variants of the graphic symbol are presented to participants in random order. Participants write down their potential interpretations of each variant. Then these interpretations are classified by three independently working judges into one of several categories, ranging from ‘correct understanding of the symbol is certain’ to ‘no response is given’ (Tijus *et al.*, 2007). This method is used in order to calculate a score for every variant. When the scores have been calculated, the results dictate which variant of a symbol is chosen based on normalized values (100 being the best, 0 the worst) (ISO9186-1:2014, 2015). When judges do not agree on a specific category, a category assigned by most judges is selected. For a symbol to pass the test, at least 67% of the participants surveyed must unequivocally understand its intended purpose in the absence of supplementary information. This metric should provide a strong indication as to whether a symbol is designed consistently with the expectations of the tested participants; however, as the test considers the view of most judges when they disagree on a specific category, the reliability of the test can be questioned. A conservative approach could be taken instead, insisting on all the judges agreeing on a category or disqualifying the symbol. Similarly, requiring consistency between different participants’ interpretations of a specific variant could make this test more reliable; additionally, one might standardize the participants’ backgrounds to ensure representativeness.

The testing undertaken during this ISO 9186 process is in line with the aims of HCI theory-based evaluation principles in terms of assessing the quality of interfaces and interactions regarding specific domains, such as public information, pharmaceuticals, and road signs (Tijus *et al.*, 2007). Moreover, it underscores the readability and comprehensibility of pictograms. Pictogram comprehension tests have been conducted independently on pictograms presented both within and outside their context. Pictograms presented within their contexts were interpreted more accurately than those shown outside them (Tijus *et al.*, 2007). Wolff and Wogalter (1998) have recommended that a pictogram should be presented in association with its context (environment) to help reduce the potential polysemy of the pictogram.

3.4 Similarity of Icons

Icons in UIs are complex elements with both visual and semantic dimensions. Designing similar icons and assessing the similarity between icons involves both visual and semantic perspectives. In this section, some methods for assessing the visual and semantic similarity of icons are presented.

3.4.1 Measuring Visual Similarity

One dimension of icons is their visual representation. Measuring the similarity or difference between icons is a long-standing problem in scientific research. There is no simple numeric measure that precisely identifies the visual features or properties of two icons that make them similar or dissimilar. While icons that are identical or share no common features at all can be readily classified, the in-between cases where some aspects are identical or show only minor differences, where elements are present in one but not the other icon, or where the size, colour, or other visual properties of the icons vary are much more complex.

There are, however, many approaches to help identify similarities between objects. The similarity between two objects is considered and measured by the number of common features shared between the objects (Estes, 1972; Reed, 2013). Visual similarity is equal to the number of characteristics shared between two objects (Taylor, 1995). The Gestalt approach discussed earlier has also focused on the relationships among and between various forms of images. In that case, the degree of similarity is identified by applying criteria to transform an object from one form to another (Singer and Lappin, 1976).

3.4.2 Measuring the Semantic Similarity of Words and Sentences: Methods

The second dimension of icons is semantic meaning. Semantic similarity has been addressed by many researchers, and many methods have been proposed for measuring semantic similarity across different applications (Budanitsky, 2006; Slimani, 2013). Semantic similarity is a concept for obtaining a metric which defines how the words themselves or words in sentences are related. It focuses on the likeness of the words' meaning or semantic content, as opposed to similarity which focuses on their likeness regarding their string format. The role of semantic similarity in this thesis is to control the semantic meanings of icons. It allows us to quantitatively compare the semantics of a source icon and generated icons which vary visually from it to a different extent.

Slimani (2013) has discussed existing semantic similarity methods based on structure, information content, and feature approaches. Furthermore, in the same publication Slimani (2013) presents several semantic-similarity ontologies. Slimani (2013) discusses three general-purpose ontologies: Wordnet, SENSUS, and Cyc KB. Slimani describes Wordnet as follows: 'Wordnet is a lexical reference system developed at Princeton University with the attempt to model the lexical knowledge of a native speaker of English. It is an online database including nouns, verbs, adjectives and adverbs grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Wordnet can be used to compute the similarity score' (Slimani, 2013). The semantic similarity score, or simply similarity score, is a value between 0 and 1 which shows to what extent two sentences are similar; 1 means that the sentences are identical, and 0 means that the sentences are dissimilar. The value depends on the algorithm that is used to semantically measure the similarity between two texts.

Wordnet was developed as a lexical system by Pedersen *et al.* (2004) at Princeton University and was the basis for many studies that proposed algorithms for measuring semantic similarities word-to-word and sentence-to-sentence, such as Shortest Path (Rada *et al.*, 1989); edge-counting (Slimani *et al.*, 2006); greedy-pairing (Lintean and Rus, 2012); and Lin's approach (Lin, 1993). Wordnet was originally created to compare words, much like one might use a thesaurus. In recent years, however, with advances in technology, it has become important to also compare semantic similarity between sentences. Many scholars report using Wordnet to compute semantic similarity between sentences (Leacock and Chodorow, 1998; Li *et al.*, 2006; Simpson and Dao, 2005).

Many of the abovementioned resources rely on different programs and require various extensions and expansions; however, Hideki Shima (2013) has released a demo of WS4J, which provides a pure Java API for several published semantic relatedness/similarity algorithms. WS4J calculates a semantic score for each of the eight algorithms, which are explained in the WS4J article in the Google Code Archive. For example, the LESK algorithm proposes that the relatedness between two words is proportional to the overlap between the dictionary definitions of these words (Banerjee, 2002). Furthermore, WS4J includes Wordnet and can be used directly through the Wordnet website. SEMILAR is a toolkit that has been implemented to measure the semantic similarity of texts (Rus *et al.*, 2013). It was developed using several algorithms that have been proposed in previous studies to measure the semantic similarity between two sentences. Greedy pairing (or greedy matching) is a linear matching algorithm that builds a solution piece by piece by selecting locally optimal solutions, followed by the algorithm choosing the best immediate solution or answer. By choosing locally optimal solutions, the greedy algorithm may also lead to a globally optimal solution (Curtis, 2003).

Lintean and Rus (2012) developed the greedy pairing approach to assess the semantic similarity between two texts by comparing the similarity between a pair of words. Initially, the algorithm constructs a set of exclusive pairs of similar words from the two texts. Then, these pairs are used to measure the overall similarity of the two texts using a weighted sum. Finally, the computed sum is optimized with the length of the texts (Lintean and Rus, 2012).

3.5 Eye-Tracking

This section provides a review of previous work using eye-tracking to collect data that helps to assess users' understanding of icons.

3.5.1 Human Vision and How Eye-Tracking Works

The movement of the human eye consists of a series of saccades and fixations. A fixation is a pause in the movement of the eye on a particular area in the field of vision. Saccades are rapid eye movements occurring between fixations that enable the generation of a complete scene viewed by a person (Bergstrom and Schall, 2014). Eye-tracking is the process of registering gaze locations and eye motions (Jabeen, 2010).

The eye's primary attention or focus is registered within the field of foveal vision. The fovea is the centre of the retina which allows the sharpest vision. When people move their eyes, they enable the foveae to register what they are looking at within the virtual line-of-sight clearly. Eye trackers are only able to track what foveal vision registers, which represents less than 1% of the human field of vision (Tobii Technology, 2012).

Fixation duration is challenging to interpret. The accuracy of participants' self-reporting of what they were paying attention to is unreliable. This is predominantly attributed to humans' imperfect short-term memory and the involuntariness of eye movements. As the capacity of working memory is finite, it limits participants' ability to remember the totality of an interaction. It appears that judgements regarding fixation duration are often unintentionally based on memory subsets that are incomplete (Wiswede *et al.*, 2007).

3.5.2 Human Vision and How Eye-Tracking Works: Prior Research

In an eye-tracking study, Guan *et al.* (2006) measured the degree to which participants failed to report elements they had viewed. It was found that in half of the cases the participants did not discuss the elements that they had looked at. To gain insight into inner emotional states and cognitive processes, researchers often need to draw on participants' subjective judgements and memory, employing think-aloud protocols, questionnaires, and interviews. Evidence from cognitive neuroscience suggests that the neural activity of an individual is often unrelated to the subjective perception they have of their behaviour (Kretzschmar *et al.*, 2013). In other words, humans are not necessarily aware of what happens in their minds. The mind-eye hypothesis states that what people see and what they are thinking is the same and that fixation equals attention (Jabeen, 2010).

Bergstrom and Schall (2014) have stressed that eye-tracking is highly effective in showing the way Gestalt design principles affect the order of the features being viewed. Eye-tracking research notes quantitative and objective measurements of eye movements as they occur, showing, for instance, shorter fixation times when reading normal text than when reading text that includes transposed letters (Liversedge and Blythe, 2007). Gaze plots can generate a visual portrayal saccades and fixations during a specific time period. In the majority of software programs, saccades are represented by lines connecting dots that represent fixations (Bergstrom and Schall, 2014).

3.5.3 Eye-Tracking and Cognitive Processes: Introduction

The research conducted for this thesis draws on related methods and techniques used to study the understanding of icons in UIs or the similarity between icons, and on protocols that have utilised eye-tracking methodology. Prior research has shown that eye fixations can provide a link between perception and cognition and that this link can be investigated using precise experimental methods, such as eye-tracking (Boeriis and Holsanova, 2012). The tracking of eye movements can be done in real-time and in a non-invasive manner to continuously monitor visual and cognitive processing (Griffin *et al.*, 2004).

3.5.4 Eye-Tracking and Cognitive Processes: Prior Research

Prior research has utilised eye movements to investigate an individual's cognitive processes and interests, particularly when viewing pictures (Buswell, 1935; Yarbus, 1967). Griffin *et al.* (2004) note that new technology has enabled eye movements to offer a sensitive, discreet, real-time index of behaviours in relation to continuing cognitive and visual processing. This progress provides evidence for the benefits of using eye-tracking in the current thesis as a method for studying working memory.

Data on eye movements derived from picture-viewing, combined with concurrent spoken picture descriptions, can provide clear indications of the dynamics of core cognitive processes (Holsanova, 2011). The integration of both eye movement protocols and spoken language descriptions can provide insight into the way an object was perceived and what drew an individual's interest and attention to parts of an image (Holsanova, 2011). Holsanova (2001) has asserted that a suitable format for integrating and analyzing both verbal and visual data is a multimodal time-coded score sheet. Over a period, the multimodal score sheet facilitates the synchronization of both verbal and visual behaviour. Holsanova (2011) has shown that this allows the content of the attentional focus to be tracked, and for clusters of verbal and visual flow to be identified.

These suggestions emerge from Holsanova's work, wherein participants participated in a two-stage experiment. In the first stage – the viewing stage – the participants viewed a complex picture. Next, during the description stage, the participants were asked to describe the picture in their own words from memory (while in front of a white screen). Eye movements were recorded during both stages. Additionally, verbal descriptions were transcribed to enable an analysis of which elements of the picture were referenced and when (Holsanova, 2011). The oral descriptions of the participants included elements such as mental state, subjective content, and the quality of their experience. Holsanova (2011) reports that the subjects disclosed compositional aspects, attributes, size, colours, referents, events, and states. The study also showed that the participants made comparisons between picture elements and created mental groupings sorting the objects into more conceptual entities, as one would expect based on Gestalt principles (Holsanova, 2011).

3.5.5 Eye-Tracking and Cognitive Processes: Conclusion

In summary, Holsanova's work shows that factors such as mental state, previous knowledge, subjective content, and the quality of experience can impact what individuals choose to focus on when viewing images. These factors can also influence viewers' subsequent recall and verbal descriptions of images. Holsanova's study is important because it shows that more information can be gathered about a user's experience and understanding by combining both eye-tracking data and verbal responses. Building on this realization, the work conducted for this thesis included both tracking participants' eye movements and asking them to provide a verbal response regarding their interpretation of the meanings of various icons.

3.5.6 Attention and Working Memory

In research literature, eye movement data is often used to explore the underlying dynamics of the cognitive processes of attention and working memory (WM), along with their interrelationship. While researchers are still determining exactly how attention and WM work together, they agree that the two do work together. Fougnie (2008) has suggested that attention and working memory interact closely during encoding and manipulation; however, attention may play a limited role in the maintenance of information. Additionally, only central attention seems to be necessary for manipulating information in working memory. The following paragraphs discuss attention and its role in WM.

Attention is a complex construct: research has shown that it can affect both the early and late stages of visual processing. In the early stages, attention can be involved in the initial feed-forward processing which occurs in the sensory cortex. For example, knowledge about the location that a visual stimulus will appear in can direct attention to this location in the early stages of visual processing (Mangun and Hillyard, 1991). In the late stages of processing, attention can be involved in the feedback processing which occurs in higher-level cognitive areas. This occurs, for example, when a participant needs to select and initiate a response (Osman and Moore, 1993).

Whether attention will modulate the early or late stages of processing depends on the demands imposed by the task (Vogel, *et al.*, 2005). Tasks with large perceptual demands may involve attentional modulation during early sensory processing, while tasks with minimal attentional demands may involve the selection of attended information only during the late stages of processing (Fougnie, 2008). One type of attention which can be influenced by perceptual demands is visual-spatial attention, a form of visual attention that involves directing attention to a location in space. Visual-spatial attention allows users to selectively process visual information through the prioritization of an area within the visual field. This type of attention is important for the studies described in this thesis, as users were asked to view images during them. Furthermore, Fougnie (2008) has found that visuo-spatial attention becomes active only when working memory stores spatial locations.

3.5.6.1 Attention and Working Memory: The Recency Effect

The recency effect is a principle which states that the most recently presented items are remembered better. This effect is relevant here as it may play a role in users' recognition of icons. Research by Wedel and Pieters (2000) has shown that as the number of fixations on a picture or brand increase, so does brand memory; however, this effect was not present when the number of fixations on text data was investigated. Furthermore, the authors found a systematic recency effect; when participants were exposed to an ad, they were later better at identifying it.

3.5.6.2 Attention and Working Memory: Eye-Tracking

As can be seen from the review of prior literature, attention, especially visual-spatial attention, and WM are linked together. These processes can furthermore be studied by using eye-tracking. For example, Sunday (2014) has demonstrated that people with higher working memory capacity have fewer eye movements and fixation periods than those with lower working memory capacity. Other studies have focused on the relationship between gaze and memory. Droll and Hayhoe (2007) found that the role of working memory is minimal when relevant information is obtained through the gaze and eye movements. This strategy is sensitive to memory load, however, because the observers' ability to actively maintain information in visual memory is limited. Visualization is limited to only one area at a time and absorbs only some of the details.

Droll and Hayhoe (2007) experimented with controlling how subjects choose to select visual cues for working memory by manipulating their ability to predict the relevancy of information for a given task. It was found that as the number of the target object features that needed to be stored in working memory increased, the observers switched from relying on their working memory to redirecting their gaze to the same location as the target object(s) to recall those features. This suggests that not all information from previous fixations is stored and that objects are often re-fixated on many times throughout the course of a task (Ballard *et al.*, 1995). Highly task-specific information is extracted during different fixations targeted by the direction of the gaze. Observers may acquire different subjective information from the point of fixation due to task demands and duration, which can vary widely. Droll and Hayhoe (2007) concluded that cognitive operation may lead to some fixation, but fixation does not define complex cognitive functioning.

Gaze direction is often interpreted as the locus of visual attention, but the operations that occur during each fixation are more complicated than a 'spatial spotlight'. This theme has been elaborated by Carlei and Kerzel (2014) who investigated the effect of unilateral gaze on visuo-spatial short-term memory. They conducted a comparison of memory performance with gaze direction to the left and right, combined with variations in the vertical positioning of the stimuli. This revealed that the direction of the gaze to the lower left was associated with better performance than a gaze to the upper right quadrant, indicating that the performance of visuo-spatial short-term memory can be affected by gaze direction. Gaze research by Palmer and Ames (1992) focused on the degree to which multiple eye fixations interfere with critical memories as a performance factor. They found that short-term memory does not persist with eye movement; rather, information needs to be encoded into a limited form of short-term memory between eye fixations.

3.5.7 Studying Recognition and Recall through Eye Fixations

According to Tversky (1974), different patterns of eye fixation can be anticipated under recall and recognition scenarios, with recognition expectations prompting a focus on the picture and recall prompting fixation on label and words. Tversky's work thus supports findings on encoding differences between recall and recognition. Furthermore, Tversky (1974) found no evidence for a relationship between correct recall and the recognition of items. Successful memory retrieval was shown to be dependent on viewing fixation patterns, while recall was enhanced by the quantity of label fixations. Recall also improved with the active generation of associations between the items, in contrast to recognition, which was increased through the assimilation of details regarding an item to allow accurate discrimination (Tversky, 1974).

Loftus (1972) investigated the extent to which the recognition memory of pictures can be predicted by eye movement patterns. In order to do this, Loftus conducted three studies. In each of these studies, the participants viewed 180 pictures and were given a yes-no recognition test on all of them. Loftus then recorded the eye movements of the participants. The results of the first experiment indicated that participants fixated more on higher-valued pictures than low-valued ones; furthermore, the participants remembered the higher-valued pictures better. When the number of fixations was held constant, however, memory performance was independent of picture value.

The results of Loftus's second experiment showed that (1) when pictures are viewed for a fixed amount of time, memory performance increases with the number of fixations on the picture, (2) when the number of fixations is held constant, performance is independent of exposure time, and (3) when pictures are only viewed peripherally, participants do not remember them. In the third experiment, pictures were viewed either normally or while the participants engaged in a distracting task (counting backwards by threes). The distracting task reduced both the number of fixations on a picture and the participants' ability to remember the picture. When the number of fixations was held constant, the performance was still better for normally viewed pictures, suggesting that the distracting task was doing more to inhibit encoding than simply reducing the fixation rate.

The results of the three studies suggest that better recall of pictures is not necessarily due to increased exposure times. Instead, increased exposure times allow viewers to have more fixations on pictures, and this increased number of fixations leads to increased memory recall performance. Based on this work, Loftus (1972) suggested that a greater frequency of fixations on a picture increases the probability that the viewer finds an easily memorisable object. This object is likely to be found relatively quickly, and the following fixations are used to focus on embedding the object of interest within memory.

Bylinski *et al.* (2015) have investigated how intrinsic and extrinsic factors can affect image memorability. Some images are intrinsically more memorable due to their content or factors related to the image. Furthermore, in Bylinski *et al.*'s (2015) studies images which are intrinsically memorable led to most or all of the participants having enhanced memories of these images. This is important because it suggests that intrinsic image memorability exists across contexts, observers, and settings, while extrinsic effects can be eliminated. In regard to extrinsic factors, Bylinski *et al.* (2015) have investigated two extrinsic factors: image context and observer behaviour. They found that images that are distinct with respect to their context are remembered better. They also recorded observer behaviour through eye-tracking. From this data, the researchers learned that as the number of fixations on an image increases, participants are more likely to remember an image. Bylinski *et al.* suggest that when designing information visualizations which one wishes to be memorable, the necessary text and pictograms should be utilised; furthermore, the principle of redundancy should be utilised.

3.6 Conclusion

This chapter discussed the concepts and methods that are important for investigating how users interpret icons on web interfaces. It provided a review of previous work on how users interpret icons and on the cognitive processes that underlie a user's ability to react to and interpret icons. A review of relevant work on methods that are used to design icons, on the similarity between icons on web interfaces, and on the context in which icons are being used was provided in order to set the stage for a discussion of the topic of this thesis. After this, the semiotics approach was discussed as it was used in the early stages of the present research to recognise the relationship between a designer-intended meaning for an icon and the user's interpretation of the icon's meaning. The final section discussed eye-tracking, as an eye-tracker was used to collect the data for this thesis. The relationship of eye-tracking with attention and working memory was also discussed.

While all of the above is important, a key issue is the fact that although there has been research into the influence of icon characteristics on a user's performance, no research has yet addressed how the visual similarity of icons influences the speed with which users recognise icons and the accuracy of their interpretations, apart from the impact of the context in which icons are used on a user's performance.

Chapter 4 presents the motivations behind Study I, along with the data collected during it, an analysis of the data, and a discussion of the meaning of the results.

Chapter 4 Exploring the Impact of Presenting Ambiguous Icons in Different Orders (Study I)

4.1 Motivation and Research Questions

The main research question of this thesis is: How do users determine the meaning of visually similar and dissimilar icons? This question is investigated through three studies and eight questions. This chapter focuses on Study I, beginning with a brief summary of the motivation for Study I, along with the research question it addresses. Thereafter, the methods used in Study I are discussed. Next, the results of Study I are presented, and the meaning of these results discussed.

When users interpret icons, many things can affect their interpretation. For example, the icon characteristics, visual representations of the icons, and the order in which they are viewed can all influence a user's interpretation. Study I was an exploratory study which investigated how users interpret ambiguous icons when they are presented with icons in different orders. In Study I, ambiguous icons (which were defined as visually identical, but with different meanings) were tested in two different orders so that the impact of icon ambiguity and the order of icons on users' interpretations of icons could be explored.

When users see icons, they employ cognitive processes to interpret the icons' meanings. Study I was conducted using open questions to allow the users to provide their reasons for interpreting the icons the way they did. Study I explored and investigated the reasons the users provided as justification for their interpretations of the icons. It was beneficial to determine which reasons the users most often cited as their justification for interpreting ambiguous icons in two different orders, as this gave the researcher insight into the users' cognitive processes. This study addressed the following research questions:

- **RQ1.1:** When viewing ambiguous icons on a limited-size screen, does the order in which icons are presented to users influence how they interpret those icons?
- **RQ1.2:** What reasons do users give for their interpretations?

4.2 Methods

4.2.1 Study Design

A lab-based exploratory study was conducted to address research questions RQ1.1 and RQ1.2. A between-subjects design was used. Two groups were created; each group was asked to undertake two tasks in different orders. Icons were presented in the context of a website. Qualitative data were collected in the form of verbalisations about the meanings of icons, both before and after participants interacted with the icons. These data were collected to determine the participants' interpretations of icons and their confidence in their level of interpretation. The participants' accuracy of comprehension was then scored using a five-point accuracy scale. The two groups were compared on their levels of accuracy of interpretation for the icons. This study took place at City, University of London.

The research process of Study I is presented in Figure 4-1.

4.2.2 Participants

A convenience sample of 20 participants was used. Most participants (N = 18) were undergraduate and postgraduate students at City, University of London. However, one participant was a postgraduate student from Cambridge University and one a primary school teacher in London. The researcher recruited postgraduate and undergraduate students in public places at the City, University of London by offering them to participate in the study. The participant from Cambridge University and the schoolteacher were recruited through a friend of the researcher.

The participants were healthy adults between the ages of 18 and 32. All participants had experience with technology and had used smartphones frequently. As such, the participants were similar in terms of maturity and their basic ability to interact with a website on a mobile device.

Two groups were randomly created; each group consisted of ten participants. In this chapter, these groups are referred to as the flight-first group and the hotel-first group to reflect the order in which icons were presented to the participants. The flight-first group consisted of four females and six males, and the hotel-first group consisted of five females and five males.

4.2.3 Materials

When designers create icons, the icons have a function; it is intended that the participants will understand this meaning or function (Ferreira *et al.*, 2005). The icon designer's intention to create icons whose meaning can be understood by users is referred to as *the intentional nature of icons*. Due to the common use and intentional nature of web interfaces, Study I was conducted in the context of web interfaces. In this thesis, 'meaning' and 'function' terms of an icon are used interchangeably.

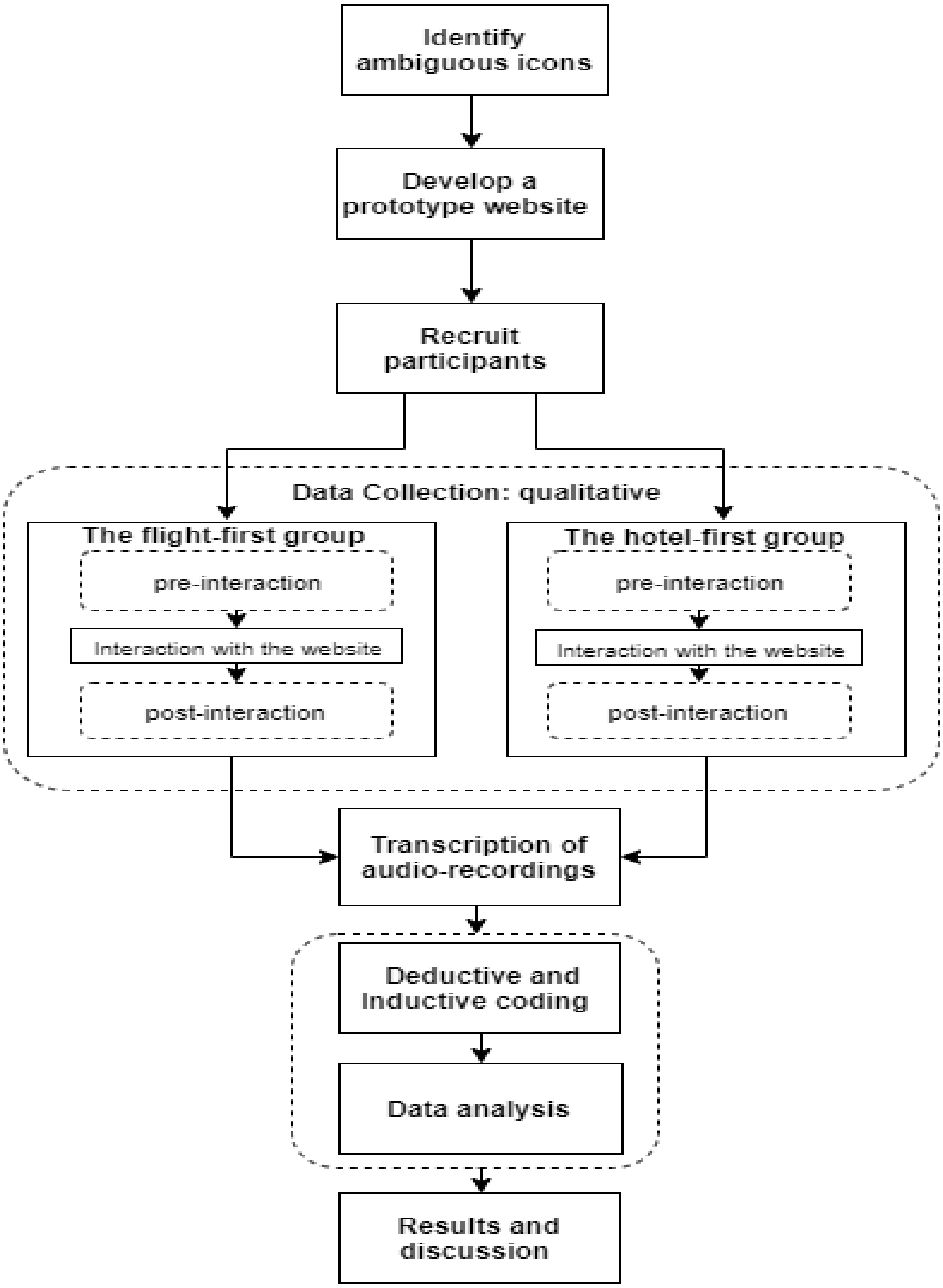


Figure 4-1: Research process of Study I.

4.2.3.1 Developing a Prototype Website

An early step in Study I involved designing and developing a prototype website that included icons. As many users have often used online websites or mobile apps to book their holidays, the icons included in the prototype website were selected after exploring and navigating many different flights and holiday booking websites and mobile apps. For instance, the websites and the mobile apps of Emirates, Qatar, Saudi, and British airlines were explored. Moreover, many online holiday booking websites were visited and navigated using a mobile device, including Booking.com, Expedia, and Trip Advisor. This exploration of potential sources allowed the identification of a wide range of icons and symbols in use, which could either be used directly in the study or after some adjustment.

Icons for the prototype website were selected for a hotel booking task (Figure 4-2) and a flight booking task (Figure 4-3). These icons were selected based on their polysemic nature. Polysemous or ambiguous icons are icons that can have more than one meaning. Icons 1 and 3 for the flight task were visually *identical* to icons 1 and 3 for the hotel task. Icon 2 was visually similar for both the flight and hotel tasks, as they were designed with an opposite pointing direction (Holiday Semiotics, 2016). Holiday semiotics (2016) is an online link to the tested website in Study I.

While Icons 1 and 3 were visually identical and Icon 2 was visually similar, the meaning of these icons was different based on whether they were in the hotel or flight task. As Study I was the first study in this research, Icon 2 in the hotel task and Icon 2 in the flight task were intentionally designed to be visually similar but not identical. This was done to explore whether the visual similarity of icons has an impact on the accuracy of users' interpretations of icons. The chosen icons were also used on other popular websites and mobile applications.

For example, the well-known sharing icon (i.e. used in WhatsApp) was used to move the current screen position to the top of the page on one of the hotel webpages (see Figure 4-2, Icon 3). It is important to note that both versions (up-pointing and down-pointing) of Icon 2 were used in Study I. Both versions of this icon were used to observe whether participants would notice a difference (up-pointing or down-pointing) between Icon 2 within a sequence.

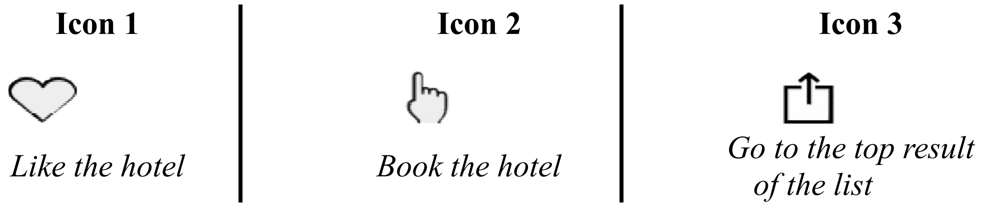


Figure 4-2: Icons in the hotel task

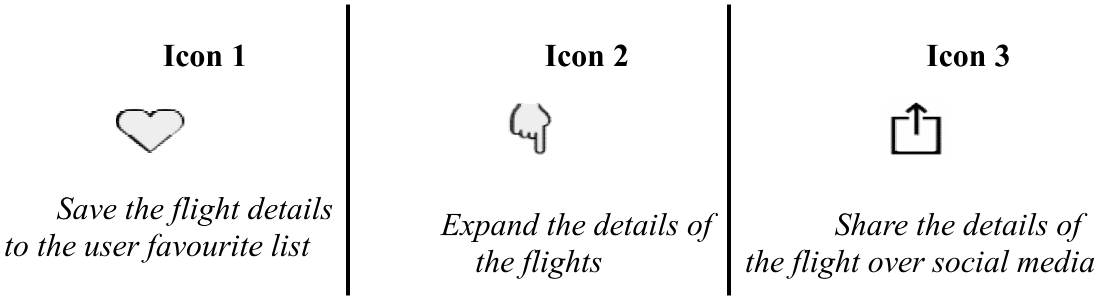


Figure 4-3: Icons in the flight task

4.2.3.2 Tasks and Procedure

All participants underwent the same procedure. However, the two groups did the tasks in a different order. The procedure lasted approximately 30 to 40 minutes. First, the participants were asked to sign an informed consent form. Before beginning the experiment, participants were asked about their level of education, the number of times that they had used holiday booking websites in the past; if they did so, they were asked about the last time they used one. The experiment consisted of the presentation of six icons, three of which were presented during the flight booking task and three during the hotel booking task (Figure 4-4). The flight booking group completed the flight booking task first, then moved on to the hotel booking task (*the flight-first group*), while the hotel booking group completed the hotel booking task first, then moved on to the flight booking task (*the hotel-first group*). Before and after the participants interacted with each of the six icons, the researcher engaged them in a semi-structured interview wherein the researcher asked questions about their interpretations and experiences with the icons (see Appendix-4.1).

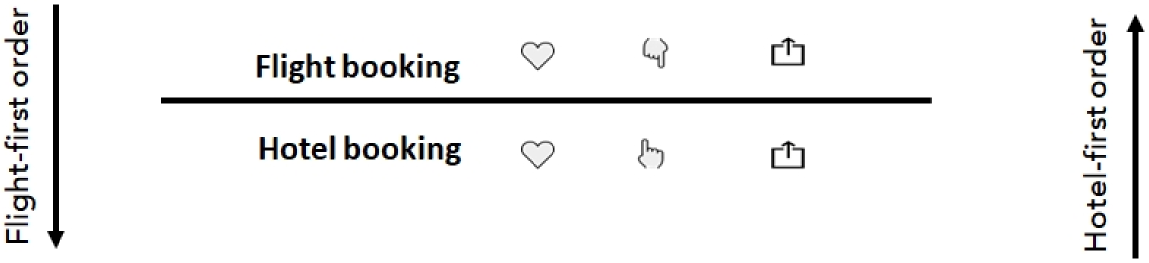


Figure 4-4: Shows the order in which the icons were presented for the flight and the hotel booking tasks. In the flight-first group, participants viewed the icons during the flight booking task then, the icons during the hotel booking task. In the hotel-first group, participants viewed the icons during the hotel booking task, then the icons during the flight booking task.

Figure 4-5 shows the steps of the flight and the hotel booking tasks for the flight-first group (For larger size screenshots see Appendix-4.2). These steps are as follows:

(1) Login

(2) Click on/touch the flight option. When clicked, it showed a flight search form that needs to be filled by the user.

(3) Click on/touch the flight search button.

(4) A list of flights is shown.

(5) Click on/touch the 'heart' icon. When clicked, a pop-up message was shown says: 'Favourite Saved'.

(6) Click on/touch the 'hand' icon. When clicked, more flight details were shown, and the relevant frame was expanded.

(7) Click on/touch the 'arrow box' icon. When clicked, a dialogue box was shown with sharing options to share the flight details through Twitter, Facebook, Google circles or e-mail.

(8) Click on/touch the home button in the corner.

(9) Click on/touch the hotel option. When clicked, it showed a hotel search form that needs to be filled by the user.

(10) Click on/touch the hotel search button.

(11) A list of hotels is shown.

(12) Click on/touch the 'heart' icon. When clicked, a pop-up message was shown says: 'Hotel has been liked by you:)'.

(13) Click on/touch the 'hand' icon. When clicked, a blank booking form was shown to complete the hotel booking process.

(14) Click on/touch the 'arrow box' icon. When clicked, the web-page was scrolled up to the top result in the list.

For more details on the meanings of all the icons, see Figure 4-6.

In the flight task, all three icons (a heart, a downward-pointing hand, and a box with an arrow) were presented simultaneously on the same screen. In the hotel task, all three icons (a heart, an upward-pointing hand, and a box with an arrow) were presented simultaneously on the same screen (Figure 4-6).

Sessions in Study I was conducted as semi-structured interviews (see Appendix-4.1). During the interview, the researcher and the participant discussed the participant's interpretation of the icons. Participants were asked to undertake the tasks and, while doing so, were asked to interpret the meaning of the icons at two points in time: *a) pre-interaction/touching:* a participant interprets the meaning of an icon before clicking on it; and *b) post-interaction/touching*: a participant interprets the meaning of an icon after clicking on it. Verbalisation was used to determine participants' confidence in their interpretations of the icons during the pre-interaction phase.

The following questions were asked to each participant to determine their comprehension of each icon. The participants were required to answer the questions aloud and their responses were audio recorded.

- During the pre-interaction phase:
    - What do you think would happen if you click on this icon?
    - Why do you think that would happen?
    - How confident are you about this icon interpretation (0% = not confident to 100% = completely confident)?
    - Do you recognize the object represented in this icon? If YES, what is that object?
    - Which part of the interface was more helpful to understand this icon's function? (icon only, icon and its context - give a percentage (%) in case of both).
- During the post-interaction phase:
    - What happened when you clicked on this icon?
    - Why do you think that happened?
    - Do you think this icon adequately represents its intended function on the interface? Why?
    - Do you have any suggestions about how this icon could more clearly represent its intended meaning?

In the post-interaction phase, in case of participants who did not get the meanings of the icons right, they were told the right meanings.

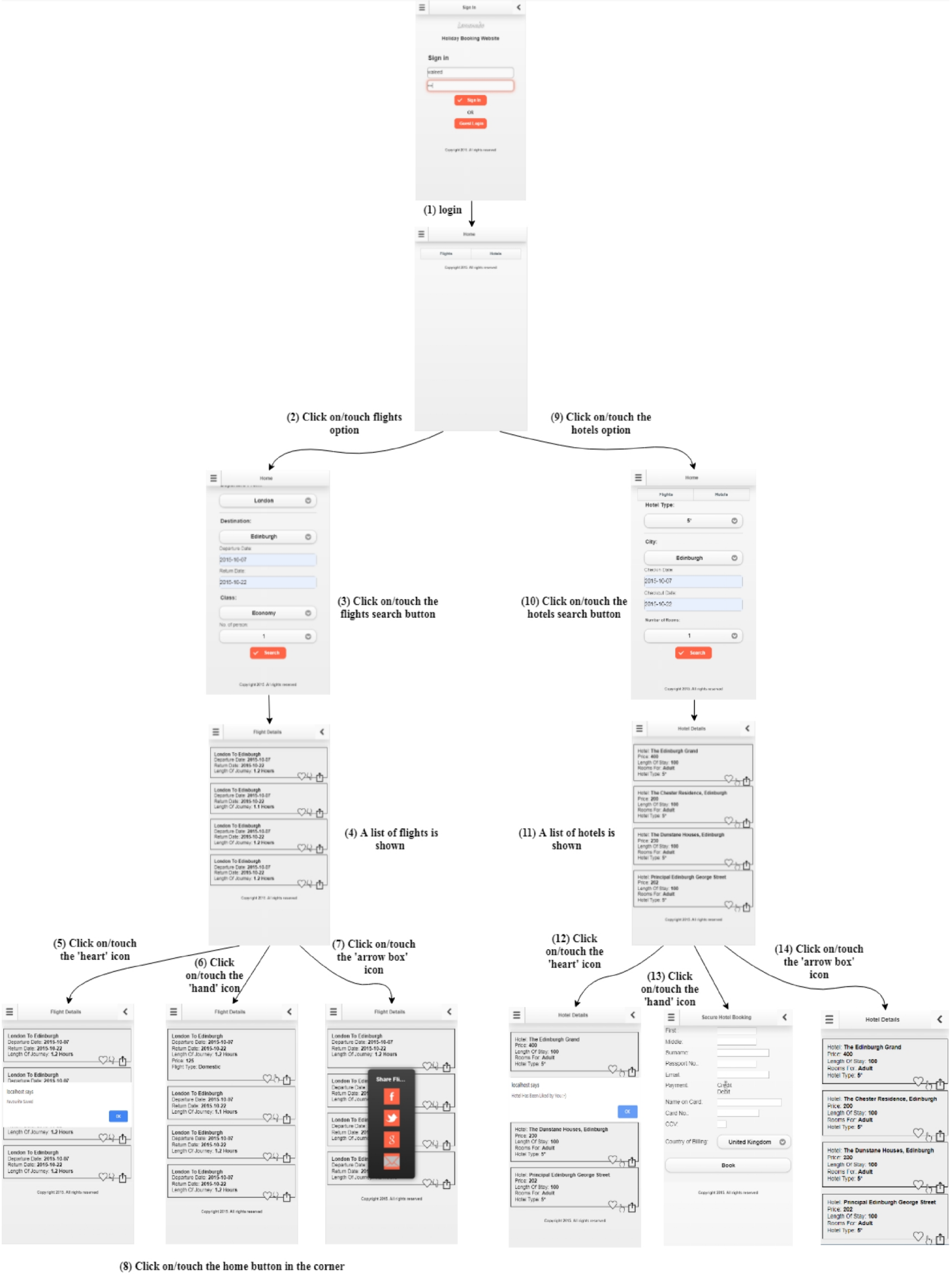


Figure 4-5: Screenshots of the hotel and the flight tasks (the steps are numbered following the flight-first order)

London To Edinburgh
Departure Date: 2015-10-07
Return Date: 2015-10-22
Length Of Journey: 1.2 Hours

London To Edinburgh
Departure Date: 2015-10-07
Return Date: 2015-10-22
Length Of Journey: 1.1 Hours

London To Edinburgh
Departure Date: 2015-10-07
Return Date: 2015-10-22
Length Of Journey: 1.2 Hours

London To Edinburgh
Departure Date: 2015-10-07
Return Date: 2015-10-22
Length Of Journey: 1.2 Hours

Copyright 2015. All rights reserved

Price: 400
Length Of Stay: 100
Rooms For: Adult
Hotel Type: 5*

Hotel: The Chester Residence, Edinburgh
Price: 200
Length Of Stay: 100
Rooms For: Adult
Hotel Type: 5*

Hotel: The Dunstane Houses, Edinburgh
Price: 230
Length Of Stay: 100
Rooms For: Adult
Hotel Type: 5*

Hotel: Principal Edinburgh George Street
Price: 202
Length Of Stay: 100
Rooms For: Adult
Hotel Type: 5*

Copyright 2015. All rights reserved

Figure 4-6: The meanings of the icons in the flight and hotel booking tasks are presented here. These options were presented after the participants searched for a flight or hotel.

4.2.4 Data Analysis

The interview data was transcribed and imported to Nvivo v11. The analysis process involved five passes of the data. In the first pass, participants' interpretations of the icons were coded according to a five-point accuracy scale utilising a deductive coding process. The second pass was performed to code the adequacy (a user judgement to evaluate how adequate visual representation of an icon represents its meaning). The third pass was performed to code the participants' confidence in their interpretation of the icons. The fourth pass consisted of finding emergent codes in the reasons provided by the participants for their interpretations; this was done using an inductive coding process. The last pass was conducted to track the changes between participants' interpretations of icons in pre- and post-interaction phases within each sequence/order. The creation and use of codes are explained further in the following sections.

4.2.4.1 Coding the Qualitative Data for Accuracy

Coding the qualitative data for accuracy was conducted in the first pass of the data analysis. Participants' interpretations of the icons were coded according to a five-point accuracy scale utilising a *deductive* coding process. The five-point scale was as follows: accurate = 5, moderate = 4, conflicting = 3, erroneous = 2, and incapable = 1. Accurate meant that the participant's interpretation matched the designed meaning. Moderate meant that the participant's interpretation matched the designed meaning to some extent, but their interpretation did not completely match the designed meaning. Conflicting meant that the participant tried to get the meaning by saying different interpretations, but in the end, they were not correct. Erroneous meant that the participant's interpretation was completely wrong. Incapable meant that the participant did not attempt to provide meaning (Islam and Bouwan, 2016).

4.2.4.2 Coding the Qualitative Data for the Adequacy

During the post-interaction phase, participants were also asked to rate the adequacy of the icons by expressing their intended meaning. The participants were asked, 'Do you think this sign adequately represents its intended function on the interface? Why?' The participants' answers were rated by the researcher on a five-point scale: very adequate = 1, adequate = 2, not sure = 3, inadequate = 4, and very inadequate = 5. A five-point adequacy scale was used to provide more information about the users' responses and to understand to what degree their responses could vary.

An example of icon adequacy is when a participant in the flight-first group was asked about the adequacy of heart icon during the flight task, they said, 'No. I don't think so, because this one means I like it; the heart shape means I like something.' However, the heart icon was used during the flight task to mean save the flight details to the user's favourite list (see Figure 4-3). Hence, this participant did not understand the meaning of the icon in the flight booking scenario and their response was coded as *inadequate* because other participants were more emphatic in their responses and they were coded as *very inadequate* (e.g., 'not at all' and 'totally'). Some responses were not covered by the adequacy scale. For example, a participant responded that 'the shape must represent the function of the icon'. This may be because the participant did not understand the question.

4.2.4.3 Coding the Qualitative Data for Confidence

In the pre-interaction phase, during the semi-structured interview, the participants were asked to report their confidence in their interpretation of the icons (see Appendix-4.1). The confidence scale ranged from 0%: not confident to 100%: completely confident.

4.2.4.4 Coding the Qualitative Data for Participant Reasoning

To further understand how participants interpreted the meaning of the ambiguous icons, it was necessary to analyse the reasons they cited for their interpretations. Inductive qualitative coding was applied to the participants' responses to the following question asked during the pre-interaction phase: Why do you think that would happen? Table 4-1 provides an overview of the 11 sub-codes that were extracted using the inductive coding process and the four main codes that were created to provide a broader overview of the reasons cited by the participants. The sub-codes provide insight into the many reasons why participants interpreted the ambiguous icons the way they did. The sub-codes are defined in Table 4-1, whereas the main codes which are labelled with 1, 2, 3, and 4 are defined in the next paragraph. Table 4-1 also provides examples of participants' responses for each of the codes (see column 'example interpretation') and shows the frequency for each reason (sub-code). While the sub-code frequency is useful, it is more important to know how many participants cited each of the main codes. Table 4-2 shows the number of participants who cited each main reason for both groups. All the participants in both groups (20 participants) cited the order of icons as the main reason for their interpretations of icons. Most of the participants (19 participants) cited the icon itself as the main reason for their interpretations of icons. Thirteen participants cited multiple contexts in the main reason for their interpretations of icons. Finally, only 11 participants cited the external context as the main reason for their interpretations of icons.

Main code	**Sub-code**	**Definition**	**Example interpretation**	**Frequency**	
				Flight-first group	**Hotel-first group**
Icon (1)	**Shape**	Refers to responses where the participants attributed the shape of the icon as the reason for their interpretations.	'It's a heart shape it symbolises love so love or liking. So, because of the shape'.	7	14
	Position	Refers to responses where the participants attributed the position/location of the icon as the reason for their interpretations.	'Because it's a finger and it has been positioned over the box and the index finger has been placed on the very top'.	4	6
	Direction	Refers to responses where the participants attributed the direction of the icon as the reason for their interpretations (i.e. pointing up or down).	'Because it looks like you know when you click something you normally get more details. and it looks like a finger pointing towards something; towards the box with the brief information so I mean I want to look at the hotel to have more details I would click on that icon to expand that.'	7	5
	Confusion	Refers to responses where the participants attributed confusion as the reason for their interpretations.	'Because this is really confusing'.	3	0

	Neighbouring effects	Refers to responses where the participants attributed the neighbouring icons as the reason for their interpretations.	'Mainly three symbols the first one that shows me the favourite of this one. This one I'm not sure but I think it going to be selecting it; it is because of the neighbour icon functionality'.	2	0
	Lack of familiarity	Refers to responses where the participants attributed the novelty of the icon as the reason for their interpretations.	'I'm looking at the icon for the first time in my life'.	2	0
	Intuitiveness	Refers to responses where the participants attributed the icon's intuitiveness as the reason for their interpretations.	'It's very intuitive that heart as a like, I understand what it means I think it doesn't require to be mention interpretation as well'.	2	0
External context (2)	**Unspecific Previous Experience**	Refers to responses where the participants attributed their knowledge or previous experience with other resources outside the study without mentioning the name of the resource as the reason for their interpretations.	'I have used it before'.	12	6
	Specific Previous Experience	Refers to responses where the participants attributed their knowledge or previous experience with specific resources outside the study, as the reason for their interpretations.	'I have seen it in WhatsApp. I have experienced this before in WhatsApp and it's for sharing'.	8	2
Order of Icons (3)	**Previous task/process**	Refers to responses where the participants attributed their interpretations to the fact that they had seen this icon when conducting the previous task or they went through the previous process or experiment.	'Mmmm. Mainly because of my previous task'.	11	8
	Previous icon/web page/page	Refers to responses where the participants attributed their interpretations to the fact that they had seen this icon on the previous page.	'Because we have used that in the previous page.'	12	16
Multiple contexts (4)	-	Refers to responses where the participants attributed their interpretations to reasons that shared more than one of these codes: icon, external context, or order of icons. The occurrence of this code is counted for this code and for the relevant individual main codes.	'Because of the icon itself; it showed me that it represents the sharing. And it's kind of that I have got the information from the last page.'	11	13

Table 4-1: Inductive coding of the reasons for participants' interpretations in the pre-interaction phase

Participants often referred to specific elements of the web-pages when they justified their interpretations of the ambiguous icons. The participants already knew some of the elements; however, there were also elements they did not know previously and only learned from the website that was used in this study, as indicated in their responses. To encapsulate the various sub-codes that were extracted from the data, three broader codes were created from the sub-codes that are shown in Table 4-1:

1. **Icon:** refers to responses where the participants attributed the icon itself as the reason for their interpretations (i.e. shape, position, or direction). Icon also refers to responses where participants attributed neighbouring icons, intuitiveness of the icons, or novelty of the icons as the reason for their interpretation. All elements that were attributed in participants' responses existed in the same/current interface.
2. **External context:** refers to responses where the participants attributed their interpretations to their previous knowledge or to their previous experience (i.e., the participants had seen the icon on other websites or mobile applications).
3. **Order of icons:** refers to responses where the participants attributed their interpretations to the fact that they had seen this icon on the previous page/previous task.
4. **Multiple contexts:** refers to responses when the participants attributed their interpretations to reasons that shared more than one of the previous codes: icon, external context, or order of icons. Multiple contexts code overlaps with other codes due to the multiple factors cited in each response.

Main code	The flight-first group	The hotel-first group	Total
Icon	10	9	19
External context	7	4	11
Order of icons	10	10	20
Multiple contexts	8	5	13

Table 4-2: The number of participants who cited each main reason for both groups.

4.2.4.5 Tracking the Change in the Accuracy of Interpretations between the Icon Pairs

To determine the impact of the order of icons on the accuracy of participants' interpretations of the ambiguous icons, pre- and post-interactions were compared. Moreover, specifically, the accuracy of participants' interpretations in the post-interaction phase for the first icons in pairs was compared with the accuracy of participants' interpretation in the pre-interaction phase for the second icons in pairs. Figure 4-7 shows the process in which the comparison between the pairs of icons was conducted.

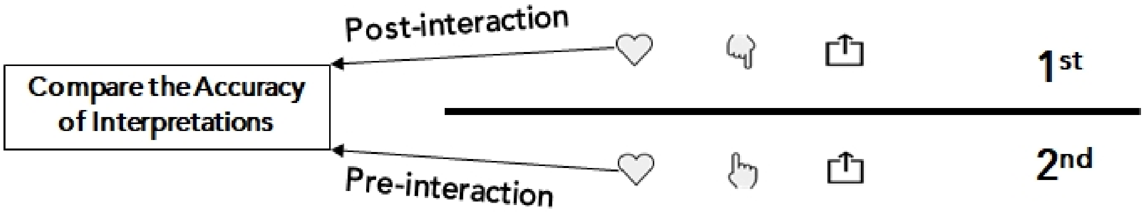


Figure 4-7: How the comparison process was conducted between the accuracy of participants' interpretations for icon pairs.

4.3 Results

4.3.1 Accuracy of Interpretation

The primary goal of this study is to understand how different presentation orders of icons influence the participants' ability to accurately interpret ambiguous icons. As discussed earlier, the participants' interpretations of the icons were coded on a five-point accuracy scale (accurate, moderate, conflicting, erroneous, and incapable). Table 4-3 and Table 4-4 show the frequency of each accuracy level for each group when participants interpreted each icon in the pre- and post-interaction phases. In this section, the pre-interaction and post-interaction interpretations of the icon pairs were analysed.

	Flight-first group						**Hotel-first group**				
Icon	**Accurate**	**Moderate**	**Conflicting**	**Erroneous**	**Incapable**	**Icon**	**Accurate**	**Moderate**	**Conflicting**	**Erroneous**	**Incapable**
♡	3	0	1	6	0	♡	0	2	1	6	1
👇	5	1	1	1	2	☝	1	1	0	5	3
⍐	4	2	1	3	0	⍐	1	1	1	7	0
♡	0	1	0	9	0	♡	2	1	0	7	0
☝	2	0	0	8	0	👇	1	0	1	8	0
⍐	0	0	0	10	0	⍐	2	0	0	8	0

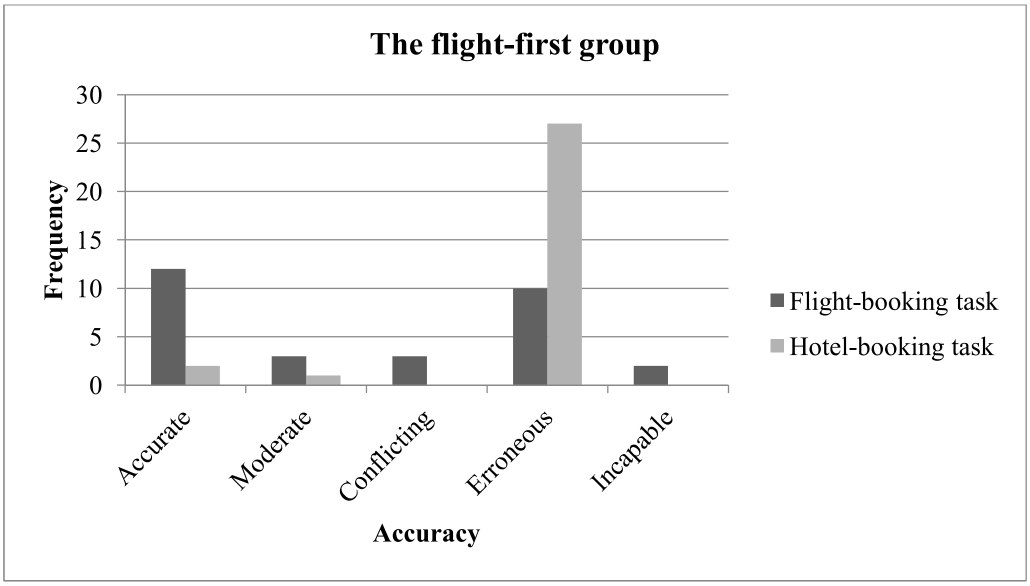
Table 4-3: The frequency of the accuracy levels in the pre-interaction phase for each group.

The frequency of accuracy in both groups during the pre-interaction phase was compared. The results indicate that the two groups were notably different in their accuracy when interpreting the ambiguous icons in the pre-interaction phase. As Figure 4-8 indicates, the flight-first group accurately interpreted the icons twice as often as the hotel-first group. Furthermore, the hotel-first group was twice as likely to be incapable of interpreting the icons correctly, compared to the flight-first group. Additionally, the hotel-first and the flight-first groups were almost equally able to interpret the icons 'moderately'. The groups were not remarkably different in the number of icons that were interpreted conflictingly or erroneously.

	Flight-first group						Hotel-first group				
Icon	Accurate	Moderate	Conflicting	Erroneous	Incapable	Icon	Accurate	Moderate	Conflicting	Erroneous	Incapable
♡	10	0	0	0	0	♡	10	0	0	0	0
👇	10	0	0	0	0	👆	10	0	0	0	0
⇪	10	0	0	0	0	⇪	9	1	0	0	0
♡	10	0	0	0	0	♡	10	0	0	0	0
👆	5	4	1	0	0	👇	5	5	0	0	0
⇪	5	3	1	1	0	⇪	5	3	2	0	0

Table 4-4: The frequency of the accuracy levels in the post-interaction phase for each group.

In the post-interaction phase, the results indicate that in most responses, the participant's interpretation completely matched the designer's assigned meaning (see Figure 4-9). When comparing the participants' accuracy pre-interaction (Figure 4-8) and post-interaction (Figure 4-9), the participants are more accurate at determining the functions of icons in the post-interaction phase. This is not surprising given that in the pre-interaction phase the icons were not supported with texts; they were presented to the participants alone. However, in the post-interaction phase, the participants had seen what the icons do in the UI, which likely helped them to accurately recognise the functions of the icons.



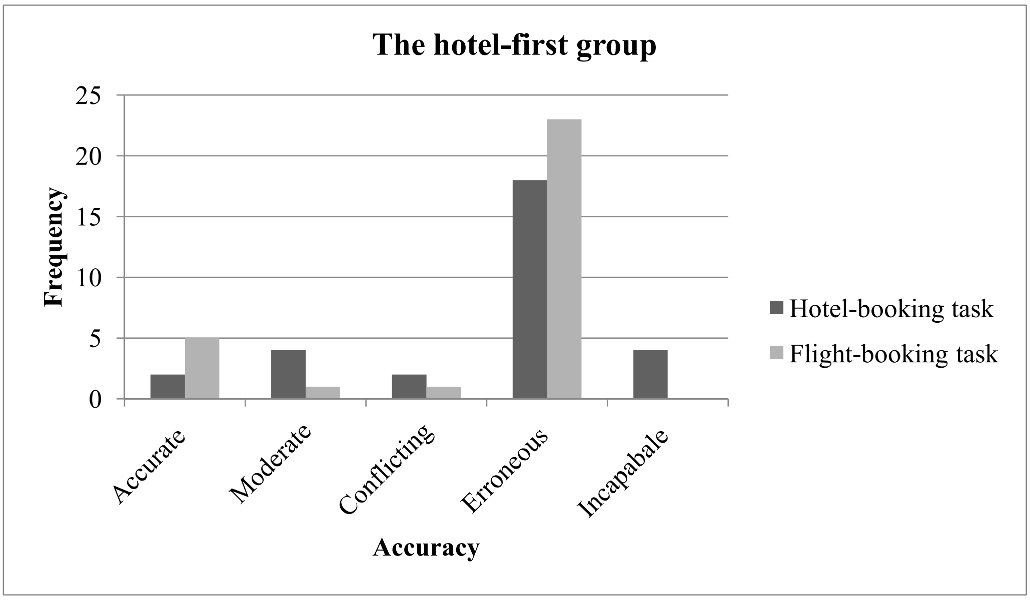


Figure 4-8: The frequency of accuracy levels for participants' interpretations in the pre-interaction phase for all the icons in each group.

Figure 4-8 and Figure 4-9 show the number of accurate, moderate, conflicting, erroneous, and incapable interpretations for all icons in each group according to the conducted task. There were ten participants in each group, and six icons were shown to each participant; that is 6*10 = 60 potential occurrences. This number is represented on the y-axis in the graph without any averaging as a simple frequency parameter.

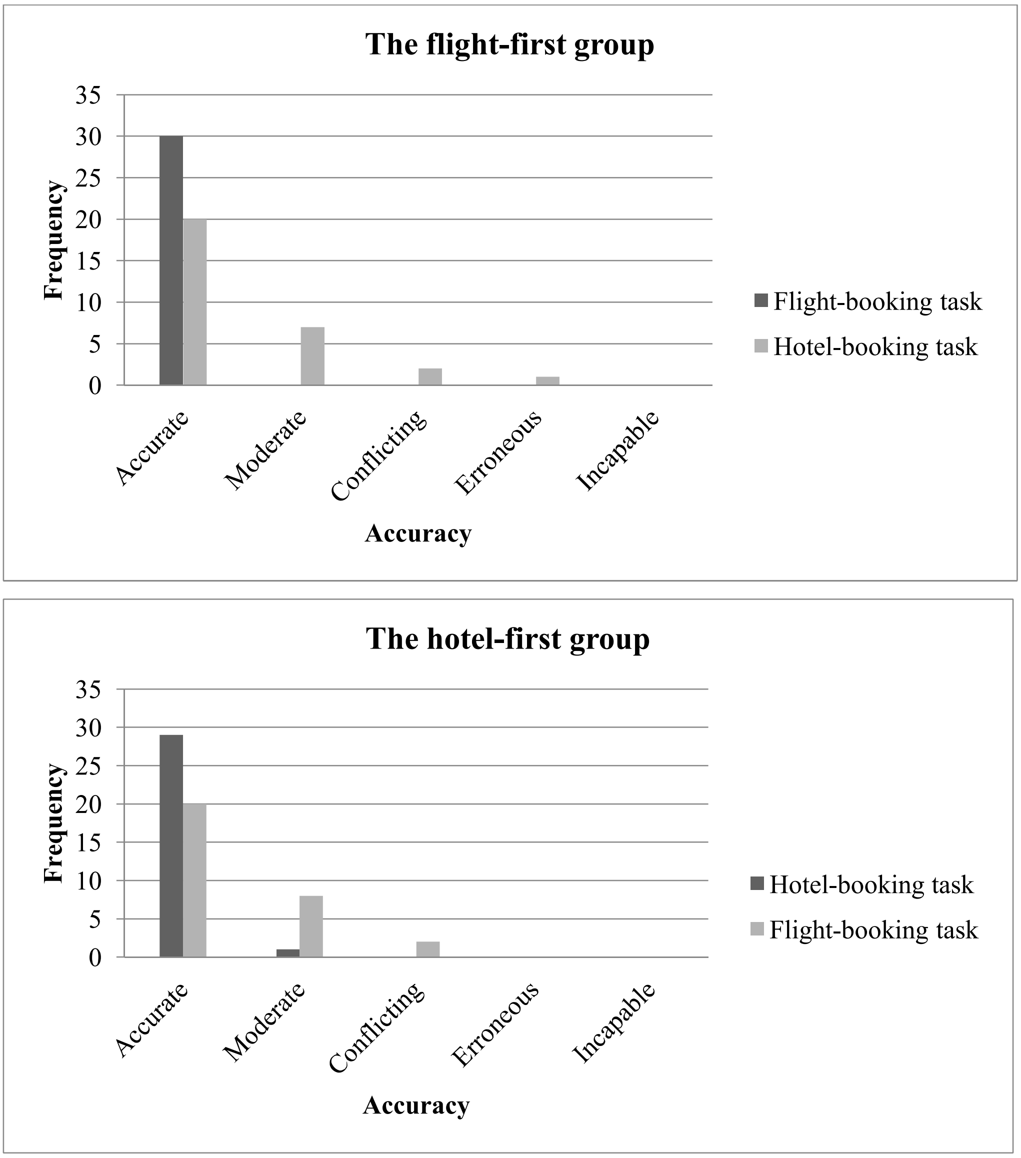


Figure 4-9: The frequency of accuracy levels for participants' interpretations in the post-interaction phase for all the icons in each group.

In most cases, for both groups, *icons and their context* were more helpful for participants than *icons only* to interpret the meaning of the icons during the pre-interaction phase, as shown in Figure 4-10. During the interview, for each icon the participants were asked, 'Which part of the interface was more helpful to understand this icon's function, the icon only or the icon and its context?' For example, the participants responded that 'the upward-pointing hand icon and its context where it appeared in the interface were together helpful to interpret the meaning of the icon'. The context refers to the web interface where the icons presented. Some participants responded that neither the icon nor the context were helpful.

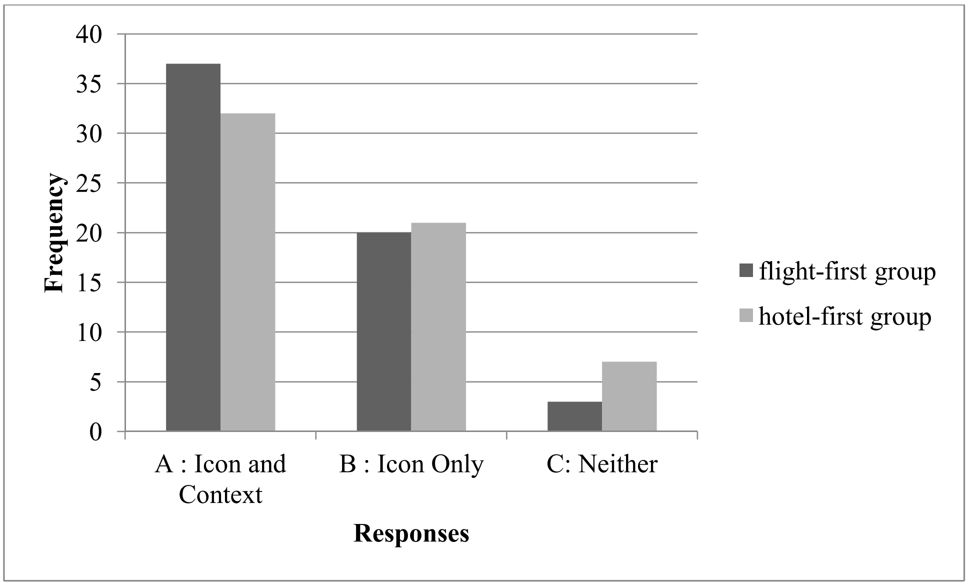


Figure 4-10: Parts of the interface that helped the participants in interpreting the meanings of the icons.

4.3.2 Adequacy of Icons' Representations

The level of adequacy perceived by participants post-interaction was scored using a five-point scale (very adequate: 1 to very inadequate: 5). The results indicate the same pattern of responses across the scale with one exception. For the flight-first group, more than double the number of icons (14 times) were perceived to be *very adequate*. Figure 4-11 shows the occurrence of the adequacy scale of the icons' representations in each group.

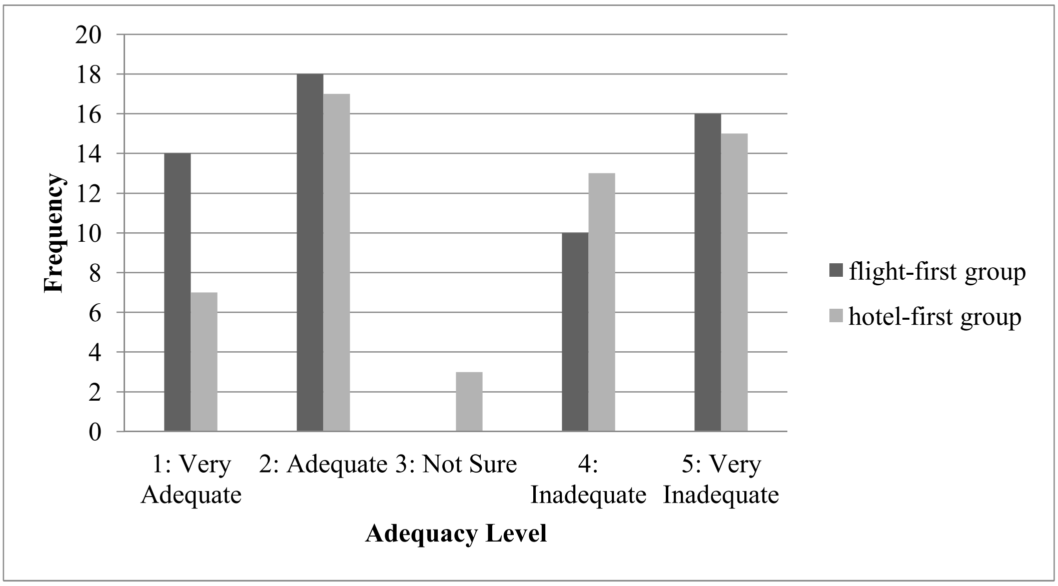


Figure 4-11: The frequency of adequacy levels of the icons' representations that were perceived for each group.

4.3.3 Confidence in Interpretation

A higher level of confidence was shown in terms of icon interpretation by participants in the flight-first group than the hotel-first group. In the flight-first group, the confidence range more than 75% and less than or equal 100% was cited 33 times compared to 27 times in the hotel-first group for all icons. Figure 4-12 and Figure 4-13 show the frequency of confidence ranges for each icon for each group.

Furthermore, in the pre-interaction phase, the participants' confidence in icon identification was compared between the first time they saw an icon and the second time they saw an icon. The results from this analysis show that participants were more confident in identifying an icon's meaning the second time they saw it. As in Figures 4-12 and Figure 4-13, the results show that for Icons 1 and 4 (heart icons), confidence levels increased from the first to the second time the icon was viewed in both groups. For the flight-first group, their confidence increased from the first time they saw the box icon (Icon 3) to the second time they saw it (Icon 6) and those who were more than 75% and less than or equal 100% confident about their interpretations nearly doubled. Table 4-5 shows detailed data sets of participants' confidences.

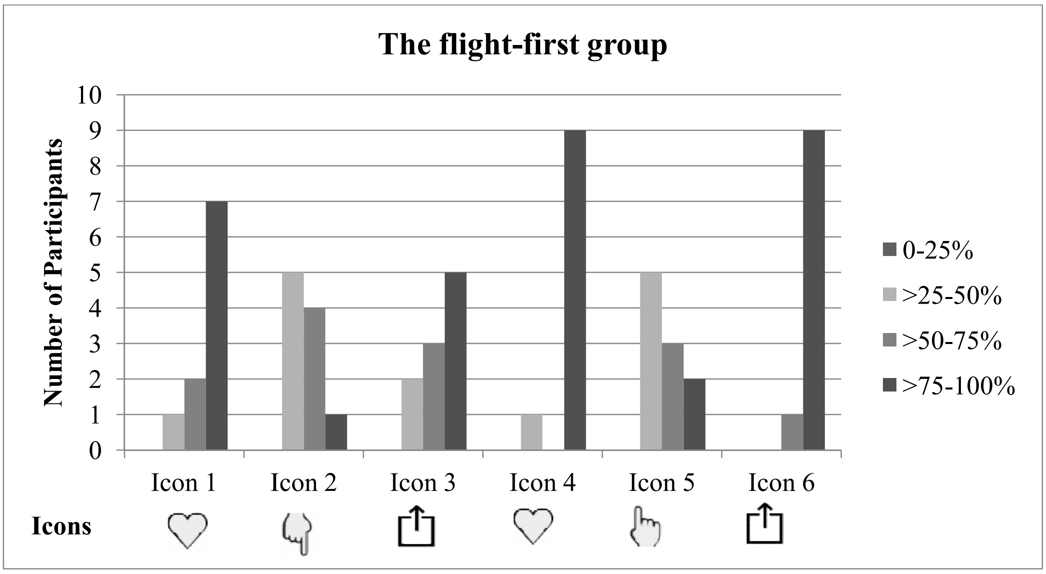


Figure 4-12: The frequency of confidence ranges in the flight-first group.

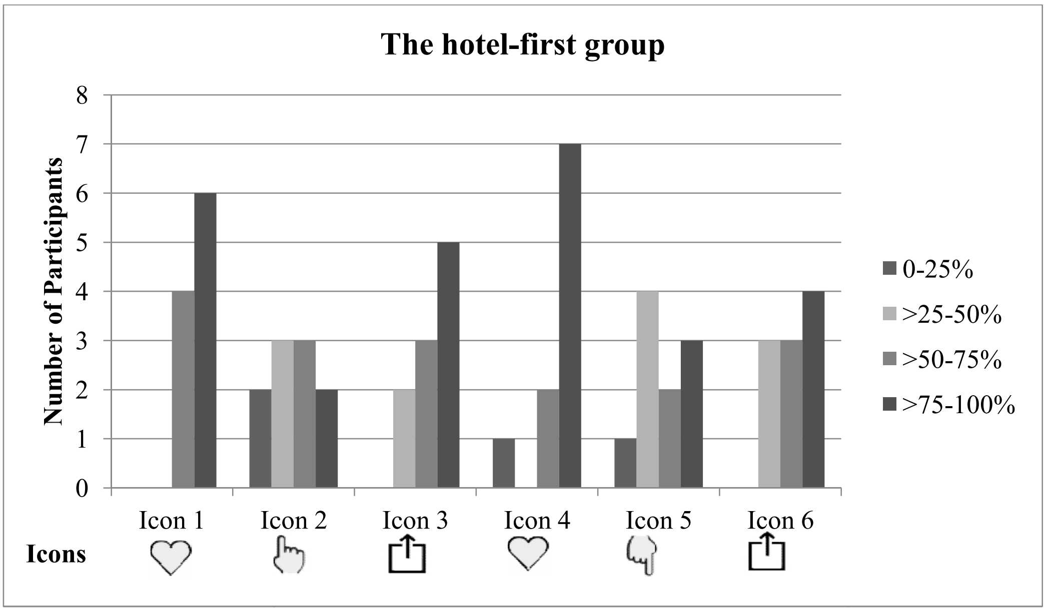


Figure 4-13: The frequency of confidence ranges in the hotel-first group.

The results presented in Table 4-5 show that for the hotel-first group, their confidence increased from the first time they saw the hand icon (Icon 2) to the second time they saw it (Icon 5), and those who were 100% confident doubled. For the flight-first group, their confidence increased from the first time they saw the hand icon (Icon 2) to the second time they saw it (Icon 5), with those who were 100% confident about their interpretation rising from 0 to 2. For the hotel-first group, their confidence increased from the first time they saw the heart icon (Icon 1) to the second time they saw it (Icon 4), and those who were 100% confident about their interpretations quadrupled.

Confidence level (%)		**0**	**5**	**20**	**30**	**35**	**40**	**50**	**60**	**65**	**70**	**75**	**80**	**90**	**95**	**100**
Flight-first group	**Icon 1**	0	0	0	0	0	0	1	2	0	0	0	4	0	0	3
	Icon 4	0	0	0	0	0	0	1	0	0	0	0	0	2	0	7
	Icon 2	0	0	0	0	0	1	4	2	1	1	0	0	1	0	0
	Icon 5	0	0	0	0	0	2	3	0	0	3	0	0	0	0	2
	Icon 3	0	0	0	0	0	1	1	1	0	1	1	1	3	0	1
	Icon 6	0	0	0	0	0	0	0	0	1	0	0	1	3	0	5
Hotel-first group	**Icon 1**	0	0	0	0	0	0	0	1	1	1	1	1	4	0	1
	Icon 4	1	0	0	0	0	0	0	1	1	0	0	0	3	0	4
	Icon 2	0	1	1	1	2	0	0	1	1	0	1	1	0	0	1
	Icon 5	1	0	0	0	0	2	2	1	0	1	0	0	1	0	2
	Icon 3	0	0	0	0	0	1	1	2	0	1	0	3	0	0	2
	Icon 6	0	0	0	0	0	0	3	2	0	1	0	1	2	0	1

Table 4-5: The frequency of confidence levels.

The results presented in Table 4-5 show that from the first time the flight-first group saw the box icon (Icon 3) to the second time they saw it (Icon 6), those who were 100% confident about their interpretation, their confidence increased quintupled. However, for the hotel-first group, their confidence decreased from the first time they saw the box icon (Icon 3) to the second time they saw it (Icon 6), and those who were 100% confident about their interpretation decreased from 2 to 1.

Confidence levels between the pairs of icons support the impact of ambiguous icons in different presentation orders. As Table 4-5 indicates, for both groups' confidence levels improved *progressively* during the tasks. When icon pairs were compared in both groups, the results indicate higher overall levels of confidence between the first time and the second time that participants see and interpret each icon.

4.3.4 Reasons for Interpretation

The primary goal of Study I was to understand the effect of different orders of icons on the accuracy of interpretations for ambiguous icons. While this analysis indicates the pattern of accuracy for ambiguous icon interpretations, further analysis is required to identify the effect of different presentation orders of icons on participants' interpretations. Therefore, the data were analysed for the second icons in each pair to identify the reasons cited by participants when interpreting the meaning of ambiguous icons, the way they did. Figure 4-14 and Figure 4-15 show the reasons cited by participants for the second icons in each pair for each group. The hotel-first group cited *order of icons* more often as their reason for interpreting the meaning of the ambiguous icons the way they did. In both groups (Figure 4-14 and Figure 4-15), the order of icons was cited by the same number of participants for Icon 5 (hand icon) and Icon 6 (box icon). For Icon 4, the order of icons was cited slightly more in the hotel-first group (nine participants) than in the flight-first group (eight participants).

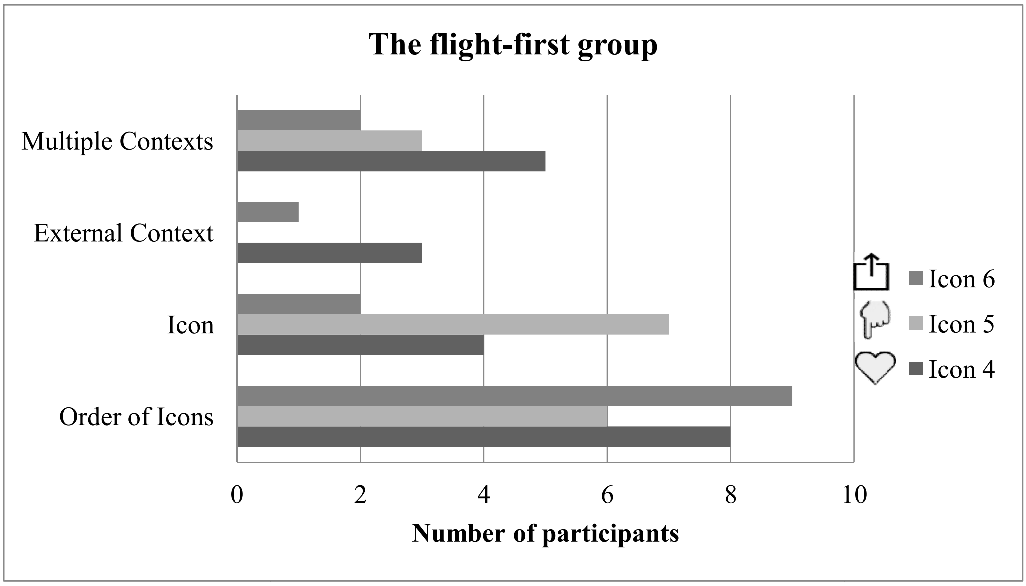


Figure 4-14: Reasons cited in the flight-first group for the second icons in each pair

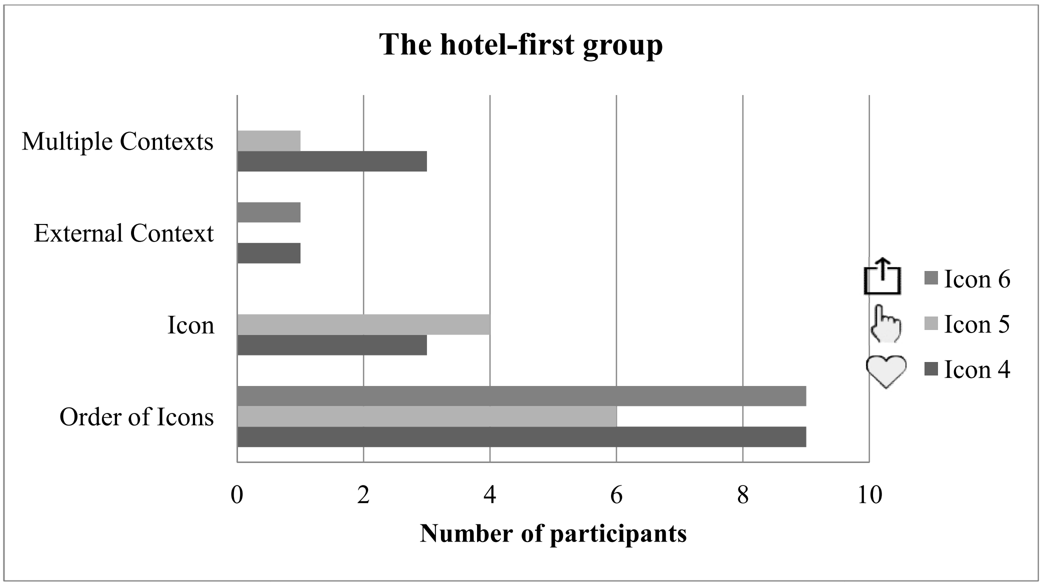


Figure 4-15: Reasons cited in the hotel-first group for the second icons in each pair

4.3.5 Comparing the Accuracy of Interpretations within each Sequence

To know how participants' interpretations of icons were influenced within each sequence, further analysis was undertaken to compare the post-interaction interpretation for one icon with the pre-interaction interpretation for the second icon in the pair. For each participant, the interpretation accuracy across the sequence is compared. Table 4-6 presents the post and pre-interactions results for the pairs of icons in the flight-first group.

User	**Icon 1 Post**	**Icon 4 Pre**	**Change**	**Icon 2 Post**	**Icon 5 Pre**	**Change**	**Icon 3 Post**	**Icon 6 Pre**	**Change**
U1	Accurate	Erroneous	-3	Accurate	Accurate	0	Accurate	Erroneous	-3
U2	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U3	Accurate	Moderate	-1	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U4	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U5	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U6	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U7	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U8	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U9	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U10	Accurate	Erroneous	-3	Accurate	Accurate	0	Accurate	Erroneous	-3

Table 4-6: Tracking the change of accuracy within the flight-first group

User	**Icon 1 Post**	**Icon 4 Pre**	**Change**	**Icon 2 Post**	**Icon 5 Pre**	**Change**	**Icon 3 Post**	**Icon 6 Pre**	**Change**
U1	Accurate	Moderate	-1	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U2	Accurate	Erroneous	-3	Accurate	Conflicting	-2	Accurate	Accurate	0
U3	Accurate	Erroneous	-3	Accurate	Accurate	0	Moderate	Erroneous	-2
U4	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U5	Accurate	Accurate	0	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U6	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U7	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U8	Accurate	Accurate	0	Accurate	Erroneous	-3	Accurate	Erroneous	-3
U9	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Accurate	0
U10	Accurate	Erroneous	-3	Accurate	Erroneous	-3	Accurate	Erroneous	-3

Table 4-7: Tracking the change of accuracy within the hotel-first group

When the participants' accuracy results and their reasons for interpretation were compared, participants' accuracy seemed to be influenced by the sequence of icons (presentation order of icons). When participants viewed an icon for the second time, they were expecting to be correct or accurate at interpreting its meaning, which they said was because they had seen the icon before.

For the flight-first group, all participants accurately interpreted the meaning of Icon 1 (heart) post-interaction, however, when they saw the heart icon again (Icon 4), most (90%) were 'erroneous' at interpreting it in the pre-interaction phase as the meanings of the icons were different to what they expected each time they saw the icons again. For the heart icon pair (Icons 1 and 4) and the box icon pair (Icons 3 and 6), the participants' accuracy decreased from the first time they saw the icons to the second. The participants tended to state that the fact that they had seen the icon before (i.e. the previous task) lead to a failure in interpreting the icons in the second time.

In the flight-first group, all the participants were 'accurate' at interpreting the meaning of Icon 2 (downward-pointing hand) in the post-interaction phase, however, when they saw Icon 5 (upward-pointing hand), most (80%) were 'erroneous', although 20% were 'accurate' at interpreting it in the pre-interaction phase. All participants were 'accurate' at interpreting the meaning of the box icon (Icon 3) in post-interaction, however, when they saw and interpreted the box icon again (Icon 6), they were all 'erroneous' (see Table 4-6).

For the hotel-first group, all of the participants were 'accurate' at interpreting the meaning of the heart icon (Icon 1) in the post-interaction phase, however, when they saw the heart icon again (Icon 4), 20% were 'accurate', 10% were moderate, and 70% were 'erroneous' at interpreting it in the pre-interaction phase. All participants were 'accurate' at interpreting the meaning of the upward-pointing hand icon (Icon 2) in the post-interaction phase, however, when they saw the downward-pointing hand icon (Icon 5), 10% were accurate, 80% were 'erroneous', and 10% were 'conflicting' at interpreting it in the pre-interaction phase. 90% of participants were 'accurate' and 10% of the participants were 'moderate' at interpreting the meaning of the box icon (Icon 3) in the post-interaction phase, however, when they saw the box icon again (Icon 6), 20% were accurate and 80% were 'erroneous' at interpreting the icon (see Table 4-7).

Table 4-7 presents the post and pre-interaction results for the icon pairs for the hotel-first group. These results reflect a similarity for icon pair (1 and 4) as those in the flight-first group in terms of interpretations and accuracy. For icons 4 and 5, the reasons cited were evenly balanced between the main codes: *order of icons* and *icon*. For Icon

6, in most cases, the order of icons was cited as the main reason for participants' interpretations.

Overall, the results indicate that ambiguity impacts remarkably on accuracy interpretations. In both groups, interpretation accuracy between the paired icons either decreased remarkably or resulted in no change. Table 4-6 shows a change in the accuracy of interpretations for more than 93% of cases (28 out of 30) in the flight-first group. The results indicate a decrease in the accuracy of interpretations between icon pairs from accurate to erroneous on the five-point scale. In most cases, participants who had 'erroneously' interpreted the icon for the first time in the pair (during pre-interaction phase) they adopted what they learned from the post-interaction phase for the same icon in the second time they saw it during the pre-interaction phase. Therefore, when the icons were 'erroneously' interpreted for the first time, the interpretation they learned during post-interaction is recalled for the same icon in the second time.

The hotel-first group had a lower percentage of cases, 83% (25 out of 30), as indicated in Table 4-7 where accuracy changed between the icon pairs. Only five cases recorded no change in the accuracy level between the icon pairs. The reason for the change is similar to the flight-first group, in that in all cases participants who had 'erroneously' interpreted the first icon in the pair (during pre-interaction phase) then adopted what they learned during the post-interaction phase for the same icon in the second time during the pre-interaction phase.

The largest effects were noted for icons 4 and 6, which were solely influenced by previous interaction events in most cases where the participants had based their interpretation in the pre-interaction phase for the second time, they had seen the icons (4 and 6). Ambiguous icons, when presented in different orders, were found to be the most effective factor in the accuracy of interpretations. The results for Icon 5 indicate that the presentation order of icons was cited less frequently. For this pair of Icons (2 and 5), a marginal difference in the representations of the icons led 70% of the participants in the flight-first group and 40% of the participants in the hotel-first group to cite the icon itself as the most frequent factor influencing their interpretation.

4.4 Discussion

This section provides a summary of the key findings of the study and presents a discussion of the theoretical and practical implications, including insights into the participants' cognitive processes. The results of Study I showed that participants' accuracy at interpreting the designer's meaning of ambiguous icons was impacted.

When the participants in this study were presented with an ambiguous icon (a visually similar or identical icon, but with a different meaning), they were inaccurate at interpreting the meaning of the icon the second time they saw it. For example, the heart icon was visually identical in both tasks (flight booking and hotel booking); however, the meaning of the heart icon was to like a hotel during the hotel booking task, while for the flight booking task, the researcher intended for the heart icon to be used to save the flight details to the user's favourite list.

The observed results may be due to the participants utilising short-term and working memory during this task. Neuropsychology differentiates various types of memory: long-term memory lasting for decades; short-term memory is only available for up to a few days; while working memory enables the consecutive and parallel steps of a complicated task to be consciously conjoined (Bouissac, 1999; Burgess and Hitch, 2005; Murphy *et al.*, 2006; Trogu, 2015). However, in this thesis, 'short-term' and 'working' memory terms are used interchangeably. Participants in this study may have drawn on working memory, consciously retaining data relevant to the given event or task (Bouissac, 1999).

The participants were presented with pairs of ambiguous icons, with a visually identical icon appearing in both a flight booking and a hotel booking task. When the participants saw the icon for the second time, they were often erroneous at interpreting its meaning in its new context, even though they did other things before seeing the second icon. These results suggest that the participants utilised short-term to remember the function of the icon over time. Hence, when they were first presented with the icon, they learned the meaning of the icon, then they used this information the next time they saw the icon. However, since the meaning of the icon had changed the next time, they saw it, they were erroneous at interpreting the meaning as they relied on their memory of the meaning of the first icon.

This study provides insights into the cognitive processes within working memory and the inherent deliberation, reasoning, and comprehension processes that influence participant interpretations (Murphy *et al.*, 2006). When a viewer interacts with a graphic, a fundamental psychological feature involved is the exceedingly short time range of working memory. In cognitive science, working memory refers to the temporary period when the mind conjoins experiences to create meaning which is then stored in long-term memory (Trogu, 2015). The significance of this dynamic was demonstrated in a study by Glanzer and Cunitz (1966) which found that free recall was associated with both short-term and long-term memory mechanisms.

The interview narratives in this study during pre-interaction and post-interaction interpretations imply a ‘sub-vocalisation’ process when participants view the icons. Within this process, the unaware, inner-speech naming of the items occurs for the objects being viewed. This is noted as a significant factor allowing people to rapidly reach meaningful closure when perceiving and understanding a visual or verbal organisation (Noizet and Pynte, 1976; Logie, 1996; Baddeley, 2014). This is consistent with Baddeley’s model (2000; 2014) that characterises the working memory component as playing a central role in binding information from two subsidiary systems: visuospatial and phonological. The model implies the interaction of both long-term and short-term memory reflected in the participants as they draw from previous knowledge and working knowledge generated from the sequence of icons presented, including icon attributes that represent sensory data within Baddeley’s model (Atkinson and Shiffrin, 1968; Burgess and Hitch, 2005).

A key piece of evidence from this study which supports the idea that the participants used short-term is that they consistently cited information from the first time they saw an icon as the most cited reason that influenced their interpretation the second time, they saw the icon. In all icon pairs, participants recalled information or interaction elements from the previous ambiguous icon. Recalling information that has been recently memorised or accessed is known as the ‘recency effect’ (Murphy *et al.*, 2006). The recency effect suggests that items that have been accessed recently are easier to remember (Burgess and Hitch, 2005).

Where the icon was viewed previously, participants immediately associated their interpretation with the previous icon. The participants’ use of working short-term memory is demonstrated by learning processes that participants describe in their interpretation and confidence levels. For example, ‘When I was in the flight’s page it was like something new; new icons; but when I transfer to the hotel’s page, I was confident with the icons given’; or ‘This icon is new for me; I wasn't expecting this before. If I face it next time, I will expect it’. This is reflected in the variation in confidence levels, and the negative change in accuracy of interpretation indicates scope to further investigate the users’ confusion induced by ambiguous icons from visual similarity perspectives between icons.

Furthermore, the pre-interaction reasons for interpretations of the second icon in each pair demonstrate the effect the order of icons to interpret the ambiguous icons. For these icons, when asking why they interpreted the icons the way they did, the participants’ responses largely fell into the *order of icons* category. Therefore, when interpreting the meaning of the second icon in each pair demonstrates the participants tended to base their interpretation on the previous icon, task, or page.

While this evidence suggests that ambiguous icons viewed in order can alter a participant's interpretations of icons, this process occurred within a broader cognitive dynamic which combines a range of contextual resources. The findings of this study also indicate that context has a major influence on users' interpretation of icons. Both groups were remarkably affected by the order of icons' presentation. Prior research has shown that most graphics constructions (e.g., icons) rely upon and are remarkably affected by the viewer's related background knowledge (Trogu, 2015). While the participants' prior knowledge may have influenced their interpretations, the responses cited in the interviews demonstrated that participants prioritised recent memory items from the previous task. However, at the same time, they did not solely depend on the information from the previous sequence interaction (for example, they also relied on the visual components of the icons).

Trogu (2015) notes that visually merging well-known past features with novel ones is the most effective way of conveying new knowledge. The interpretation process involves the viewer applying a complicated network of expectations consisting of past experiences. Thus, participants determine the meaning of icons from both external and study contexts. In modern psychology, perception and remembrance are considered productive actions rooted in previous experience (Trogu, 2015). The implication is that while internal contexts (the study context) related to the previous task in the sequence are remarkable, high familiarity and relevance of the prior knowledge to the topic matter can help to overcome the limitations of working memory (Trogu, 2015). In other words, although information about the prior icon seems to influence participants' interpretation of the second icon, it may be possible that this influence can be overridden by knowledge. Hence, although icon interpretation may be a bottom-up process, it may be possible to control this interpretation through top-down processes.

While the participants often cited sequence effects when interpreting the second icon, some other notable information came from asking the participants how they interpreted the icons. Participants cited various reasons as mentioned in Table 4-1. This suggests that the most recent items they viewed influenced contextual representations. Furthermore, participants indicated the importance of icon attributes and sensory information in constructing meaning for the icons (Atkinson and Shiffrin, 1968). Additionally, participants consistently mentioned the role of text in support of their interpretations. This implies that graphics can be misunderstood and misinterpreted without supporting textual context (Trogu, 2015). The positioning or direction of icons was cited on multiple occasions as a factor that affected icon interpretation. There is some indication of a co-occurrence/neighbouring effect in a small number of cases. Participants referred to a neighbouring icon as a co-occurrence factor in their interpretations.

Furthermore, the intuitiveness of the icons was cited by a small number of participants as an important factor in their interpretation. Maximising this factor can support the interpretation of icons, particularly for novice users who are unfamiliar with the icons. Participants also reported signification, which refers to the icons' symbols, meaning, or representation. In some cases, syntactic elements were noted in participants' interpretations such as position, clarity, and presentation of the icon (Islam and Bouwan, 2016).

Furthermore, there is an extensive reference in the participants' responses to external contexts which may be part of *specific* or *unspecific* previous experience. Participants cited a range of different external contexts from their external experiences in interpreting icons: their background knowledge is drawn from experience with social media and mobile device applications and in some cases the desktop applications. One participant acknowledged a diversity of external contexts where the same icon is used for different purposes. Some icons were recognised as *common* icons or *universal* standards, widely used in external applications and websites and associated with common functions such as 'favouriting', liking, or sharing. Participants emphasised the importance of text support, while one participant stated it reinforced their interpretation.

There is some evidence to suggest the importance of consistency between external contexts and the context of the study (Study I). Remarkable confusion was noted in participants' interpretations of icons. One participant's comment highlight multiple factor, 'the information and the context are not enough and the icon is index finger pointing up; so it's really complicated for me'. In this study (Study I), it was noted that inconsistency or conflicting signals between external context and the context of Study I generate remarkable levels of confusion that mislead participant interpretations.

Despite the rich data referring to participant external contexts, participants greatly utilised the context of the study when interpreting the icons' meanings. The results from this study provide some indication of the contextual representations emphasised by (Burgess and Hitch, 2005) in contrast to short-term memory models. Users in this model retrieve and evaluate information from contextual items. (Burgess and Hitch, 2005) suggest the relative significance of different items stated in terms of a recency-weighted sum of the context from each item. Consistency was a key theme emerging from the interviews with participants emphasising that the use of icons throughout the entire sequence should be consistent. This can minimise confusion and maximise recency effects. Further, subsequent rehearsal and retrieval can help conversion to long-term memory if the same symbols are used within the presentation order of icons.

4.5 Conclusion

In summary, the research questions presented in Study I were, RQ1.1: When viewing ambiguous icons on a limited-size screen, do different orders in which icons are presented to users influence how they interpret those icons? RQ1.2: what reasons do users give for their interpretations? Study I gave insights into the dynamics between the user and the interface on which they view icons and understanding the cognitive process of users while they interpret icons. These insights were provided by assessing the cognitive processes of the users before and after they interpreted the meaning of icons. The users' cognitive processes were probed with 'what' and 'why' questions. For example, participants were asked, 'what do you think would happen if you click on this icon?' And 'why do you think that would happen?' The 'what' and 'why' questions used in this study gave insights into the users' interpretations of icons.

Furthermore, the users were asked to report how confident they were in their interpretation. Therefore, Study I provided various types of information such as interpretations, reasoning, and confidence about the cognitive processes of the users as they interacted with and interpreted the meaning of the icons

Based on the reasons cited in the study, presenting ambiguous icons for the first time had a remarkable influence on subsequent interpretations when they are presented for the second time. This pattern is reflected for icons 6 and 4. For Icons 2 and 5 (thumb pointing upward and thumb pointing downward), where the icons were marginally different in visual representations, a couple of participants interpreted the icons 'accurately' or 'conflictingly' the second time they saw the similar icon (Icon 5). Hence, a marginal difference in visual representations of icons might affect participants' interpretations of these icons; this will be further investigated in the next study (Study II) to better understand this effect.

The results of this study indicate that when users view ambiguous icons in different orders on a limited-size screen, how they interpret those icons is influenced.

The results show that when users are provided with the meaning of an ambiguous icon and then later must interpret the meaning of a visually identical icon with a different meaning, their interpretation of the second icon is often ‘erroneous’. This pattern was not only seen in the accuracy of participants but was also reflected in their verbal responses of how they interpreted the icons. Hence, this study suggests that users’ ability to interpret ambiguous icons is influenced by context, along with working memory. These findings led to Study II, which investigated how the extent of visual similarity between icons presented in different orders on a limited-size screen can influence user interpretation of those icons.

Chapter 5 Exploring the Impact of Icon Similarity on Users' Speed of Recognition and the Accuracy of their Interpretations (Study II)

5.1 Introduction

In this section, firstly, the motivation behind Study II is discussed. Secondly, a limited number of methods that were proposed in the literature to measure the visual similarity between icons are surveyed. Thirdly, it discusses the potential effects on user performance when visually similar icons are viewed. Finally, the research questions of this thesis are presented.

5.1.1 Motivation

The results of Study I indicated that viewing ambiguous icons sequentially influences how users interpret those icons. Study II expanded on these findings by investigating the extent to which the visual similarity between icons presented in sequence can influence users' interpretation of icons. Hence, the objective of Study II was to identify the impact of visually similar icons on participants' speed of recognition and accuracy of interpretation of icons.

Users typically encounter various icons while navigating a website. Any two icons will vary in their degree of similarity (or dissimilarity). For a user who is new to a website, accurately comprehending the meaning of the icons is a key aspect of being able to use a website effectively. Where icons are readily and accurately interpreted, tasks will be completed quickly and efficiently, whereas if icons are misunderstood, there will be more errors, and tasks may take longer, or fail to be completed at all.

When a user first encounters an icon, they need to visually inspect it and reason about its context (both the visual context and the relevant context of task or process that the icon appears within), to interpret its meaning. This complex activity involves two different levels of processing: *visual* and *semantic* processing. The visual processing refers to processing the visual aspect of an icon – what it looks like; whereas the semantic processing means extracting the meaning or function of the icon based on values associated with its appearance. For example, if a user sees an icon consisting of four stars on a travel website, this might immediately create an association with reviewing, as several stars is often used as a measure of quality on travel websites.

Unsurprisingly, this complex process has been investigated before to begin to understand how users make these two associations between shape and meaning. Prior research has used eye movements to obtain a detailed account of a user's cognitive processes when viewing pictures (Buswell, 1935; Yarbus, 1967). This early research had to investigate this issue without the benefit of modern eye-tracking equipment. Thus, the focus of Buswell (1935) and Yarbus (1967)'s work was on the cognition involved in interpreting images rather than on eye movements and visual focus. One fundamental measure that emerged in that work was the time that users spend viewing the shapes. Increased time viewing an image suggests an increased cognitive effort is needed to recognise or understand an image.

The mind's eye hypothesis states that what people see and what they are thinking are directly related, and that fixation equates to attention (Jabeen, 2010). This assumption is a foundation for modern eye-tracking research. Eye fixations, where a user's attention dwells for more than a fraction of a second, are viewed as the interface between perception and cognition. That detail can then be investigated with precise experimental methods (Boeriis and Holsanova, 2012). Eye movements can facilitate real-time analysis in a non-invasive manner to monitor a user's continuous visual and cognitive processing (Griffin *et al.*, 2004). Controlled studies can capture eye movement patterns (global and local), duration of fixations, scan-paths, and spatial and chronological sequencing of fixations (Boeriis and Holsanova, 2012).

The first stage, people scanning the visual content in front of them, directly represents the first stage of visual processing. In the second stage, users work towards deciding on the icons' meanings, and in a computer interface, the functions that are associated with that meaning. Eye-tracking alone cannot reveal the meaning-making that transpires at this point in the user analysis of the scene before them. Researchers have endeavoured to uncover user behaviour by a variety of methods, but particularly through asking users to verbally describe their reasoning. Tying this verbal report to user visual behaviour requires careful analysis. Typically, the integration of both eye movement protocols and spoken language description is a key goal to enable accurate insights into what drew a user's interest and attention and how an object was perceived (Holsanova, 2011). The accuracy of how users comprehend the meaning of an icon can be captured through the precision of their interpretations of the icon's shape, meaning, and function. An exact interpretation of an icon's meaning indicates that the user has successfully understood its meaning; whereas an inaccurate interpretation indicates that the user is confused or in error about the function or meaning of the icon. Note that a user may reassess their understanding after experimentation, e.g. after clicking on a button. While their initial expectation may be in error, after seeing the system's reaction, they may come up with a revised, and more accurate, understanding of what the icon means. At this point, users may be prone to recasting their initial (inaccurate) expectation as more accurate than it was.

Thus, in assessing the people' visual interpretation of an icon, the total time they spend viewing the icon can be assessed before arriving at an explanation of its meaning. In study II, the viewing time on an icon is defined as the 'dwell time': *the total amount of time a user fixates on an icon in silence until a response from the user is indicated.* Second, the accuracy of that understanding can be assessed *before* interaction occurs. People must take care to identify their initial understanding before action is taken, as their interpretation after interaction may be more accurate, but not an honest reflection of their initial impression.

The main goal of study II was to investigate how the visual similarity of icons influences users' recognition and understanding of the icons. Prior research has shown that humans recognise familiar faces efficiently and fast. Not surprisingly, icons that have been seen previously are more quickly and effectively recognised by users than unfamiliar icons (Reber *et al.*, 1998; McDougall and Reppa, 2008). The similarity between icons plays a role in enhancing subjects' familiarity as a feeling (Moreland and Zajonc, 1982). However, the effect of the degree of icon similarity on users' speed of recognition remains largely unknown. Moreover, variable users' speed of recognition for similar and dissimilar icons may result in varying accuracy levels of their interpretation.

5.1.2 Measuring the Visual Similarity between Icons

Measuring the similarity or difference between icons is a long-standing problem in scientific research. There is no simple numeric measure that precisely identifies the visual features or properties of two icons that make them similar or dissimilar. While icons that are either identical or share no common features at all, are readily classified, the complex area between where *some* aspects are identical, *some* are nearly identical but have minor differences, where elements are present in one, but not another, or size, color or other visual properties vary is much more complex. Many methods have been used to determine the visual similarity between icons (Estes, 1972; Reed, 2013; Taylor, 1995; Singer and Lappin, 1976), however, there is no single standard method for assessing these complexities.

5.1.3 Icon Visual Similarity and User Recognition

This study aimed to understand in greater detail how users ‘make sense’ of new icons. From the previous literature, it would be expected that assessing the meaning of a known icon that fulfils it's expected (previously understood) purpose should be both exact and fast. Where an icon is visually identical to a previously-encountered, well-known icon, recognition will be fast. However, if its function is different, the user might make an error as they associate the shape with a different meaning. Conversely, a completely novel icon whose shape has not been seen before will demand extended attention and reasoning about its potential purpose. The initial interpretation may be no more than a guess based on real-world metaphors, previous experience of similar (but not identical) icons, etc.

Furthermore, it is thought that users will attempt to recognise novel icons based on their experience with icons that share visual similarities. For example, if a user encounters a thumbs-up icon on a new website, they are likely to recognise it as being like the common ‘like’ icon found on Facebook. Since the icon is visually like the Facebook icon, the user will likely presume it has the same function, i.e. to indicate their liking of some content located next to the thumbs-up icon.

Users are also influenced by recency. If they have just used Facebook, they are more likely to think that an icon which is visually like Facebook’s ‘like’ icon has the same meaning. When a user encounters a new icon, they are more likely to recognise it again a short time after their first encounter. If, on the other hand, they next encounter it weeks later, the chances of an association being made are small.

How recency and recognition function when the association happens in a particular instant is not obvious. When icons are not identical but only visually similar, the strength of the association that users will have for these icon types is currently unknown. How rapidly the effect of recency declines or how multiple possibilities interact with each other when a user encounters multiple new icon is also not clear.

In Study II, users were exposed to four variations of an icon while their eye-movements were tracked with an eye-tracker. The purpose of the icons was the same (i.e. the fundamental semantics), but the visual presentation varied. It is thought that since the experiment was short (approximately 30 minutes) any recency effects would be very short-term. The previous literature has not systematically assessed how localised time recency effects are. One reasonable hypothesis would be that in very short periods, the differences in recency effects will be minimal: across a task of approximately 15 minutes, there would be no variation in time or accuracy performance in recognising recently encountered icons. Another reasonable hypothesis is that the influence of recency declines rapidly, and accuracy and time performance in interpreting icons will vary markedly even when a task only lasts a few minutes. There is no clear guidance from the literature as to which of these competing hypotheses is more likely. However, Study I suggest that the effect of recency has its strongest impact in the first few seconds, and then rapidly decays. Therefore, within tasks of a few minutes, it would be expected that differences emerge based on the ordering of the similar sets of icons.

5.1.4 Research Questions

Study II investigated the impact of visual similarity between a known 'target' icon and icons in a series, viewing times, and the accuracy of participants' interpretation of the icons. The order of presentation was also investigated. There were two possible presentation orders; one began with an icon which was identical to the known 'target' icon and ended with an icon that was completely different/dissimilar[3] to the known 'target' icon, (the 'forward' presentation order); the other presentation order was the opposite (the 'backward' presentation order). Viewing times and accuracy of the interpretation of meaning for the icons in these two sequences were compared to investigate the influences of similarity over time. Accordingly, Study II addressed the following research questions:

- *RQ2.1: Will participants who are presented with visually similar icons in a 'forward' order recognise the icons at a different speed than participants who are presented with visually similar icons in a 'backward' order?*
- *RQ2.2: What is the impact of the degree of visual similarity between icons on the speed with which a user recognises the icons?*
- *RQ2.3: Will participants who are presented with visually similar icons in a 'forward' order interpret the icons with different accuracy than participants who are presented with visually similar icons in a 'backward' order?*
- *RQ2.4: What is the impact of the degree of visual similarity between icons on the accuracy with which a participant interprets the icons?*
- *RQ2.5: Is there a correlation between speed of recognition and accuracy of interpretation for each icon type (the four icon types)?*

---

[3] The terms 'completely different' and 'dissimilar' are used interchangeably in this thesis.

5.2 Methods

In this section, the methods for conducting Study II are presented. It discusses the design of the study, the participants, the materials that were used to conduct the study, the tool that was used to collect the eye-tracking data, how the visual and verbal data were collected, the procedure followed in running the experiment, the preparation of the data for analysis, and finally, the data analysis.

5.2.1 Study Design

A lab-based experimental study was conducted to address the research questions of Study II. The independent variables (IVs) were the visual similarity of icons ('icon type')[4] and the icon presentation order. The dependent variables (DVs) were the speed of users' recognition and the accuracy of interpretation. The visual similarity of icons had four levels: identical, slightly different, mostly different, and dissimilar. The icon presentation order had two levels: 'forward' and 'backward'. The presentation order served as a *between-subjects* factor and the icon type served as a *within-subjects* factor.

Implementing presentation orders and icon types in a single task type (block) resulted in a 2x4 Latin rectangle as an initial experimental design. However, the number of treatments was small (two treatments only) and it was thought that the experiment may involve carryover effects (recency effects) over time when measuring the DVs at different levels of the IVs. Therefore, to increase the degree of freedom for experimental error and consider the potential variation of recency effects over time, an additional task type (block) was added to the design of the experiment. All participants completed two task types where the same presentation orders for the icons were retained (the presentation order in the first task type was the same as the presentation order in the second task type). Some participants were presented with icons in 'forward' order in the first task type followed by icons in 'forward' order in the second task type. Other participants were presented with icons in 'backward' order in the first task type followed by icons in 'backward' order in the second task type. However, the additional task type may have also involved carryover effects (learning effects). Therefore, task types were counterbalanced with presentation orders to address carryover/confounding effects such as recency and learning effects. Counterbalancing task types with the repeated presentation orders together resulted in repeated 4x4 Latin squares (see Table 5-1).

---

[4] The terms 'visual similarity of icons' and 'icon type' are used interchangeably in this thesis.

As a result, the design of the experiment consisted of two task types (hotel booking and flight booking), two presentation orders ('forward' and 'backward'), and four icon types (identical, slightly different, mostly different, and dissimilar). This design resulted in four groups/treatments (1. hotel-flight ('forward'), 2. hotel-flight ('backward'), 3. flight-hotel ('forward'), and 4. flight-hotel ('backward')). Each group was presented with the four icon types (see Table 5-1). As the task type was not part of the IVs in this experiment, the data from the two task types were combined during analysis.

The hypotheses and null hypotheses for each research question in this study were as follows:

- For RQ2.1:
    - **$H_{2.1}$:** Participants who view the icons in the 'forward' presentation order will be significantly quicker at recognising the icons than participants who view the icons in the 'backward' presentation order.
    - **H_0:** The 'forward' and 'backward' presentation order groups will not recognise the icons at a significantly different speed.
- For RQ2.2:
    - **$H_{2.2}$:** The participants' speed of recognition will be significantly different between the four icon types, in which participants will recognise the icons that are identical to the known 'target' icon the quickest, participants will recognise the slightly different icons slower than the identical icons, participants will recognise the mostly different icons slower than the slightly different icons, and finally, participants will recognise the icons that are dissimilar to the known 'target' icon the slowest.
    - **H_0:** The participants' speed of recognition will not be significantly different between the four icon types.
- For RQ2.3:
    - **$H_{2.3}$:** Participants who view the icons in the 'forward' presentation order will be significantly more accurate at interpreting the meaning of the icons than participants who view the icons in the 'backward' presentation order.
    - **H_0:** The 'forward' and 'backward' presentation order groups will not be significantly different at accurately interpreting the meaning of the icons.
- For RQ2.4:

    - **$H_{2.4}$:** Participants' accuracy at interpreting the meaning of an icon will be significantly different between the four icon types, in which participants will interpret icons that are identical to the known 'target' icon with the highest accuracy scores, participants will interpret slightly different icons with lower accuracy scores than the identical icons, participants will interpret mostly different icons with lower accuracy scores than the slightly different icons, and finally, participants will interpret the icons that are dissimilar to the known 'target' icon with the lowest accuracy scores.
    - **H_0:** The participants' accuracy of interpretation will not be significantly different between the four icon types.
- For RQ2.5:
    - **$H_{2.5}$:** For each of the four icon types, there will be a significant positive correlation between the speed of recognition and accuracy of interpretation.
    - **H_0:** For each of the four icon types, there will be no significant correlation between the speed of recognition and the accuracy of interpretation.

The research process of Study II is presented in Figure 5-1.

Group	**Task1 (Block)**	**Task2 (Block)**	**Visual similarity of icons (Icons types)**			
			1	**2**	**3**	**4**
1	Hotel	Flight	Identical	Slightly different	Mostly different	Dissimilar
2	Hotel	Flight	Identical	Dissimilar	Mostly different	Slightly different
3	Flight	Hotel	Identical	Slightly different	Mostly different	Dissimilar
4	Flight	Hotel	Identical	Dissimilar	Mostly different	Slightly different

Table 5-1: shows the order in which participants conducted the hotel and the flight tasks (Task 1 and Task 2) and shows the order in which participants viewed the icons (Groups 1 and 3 received the 'forward' presentation order, while Groups 2 and 4 received the 'backward' presentation order).

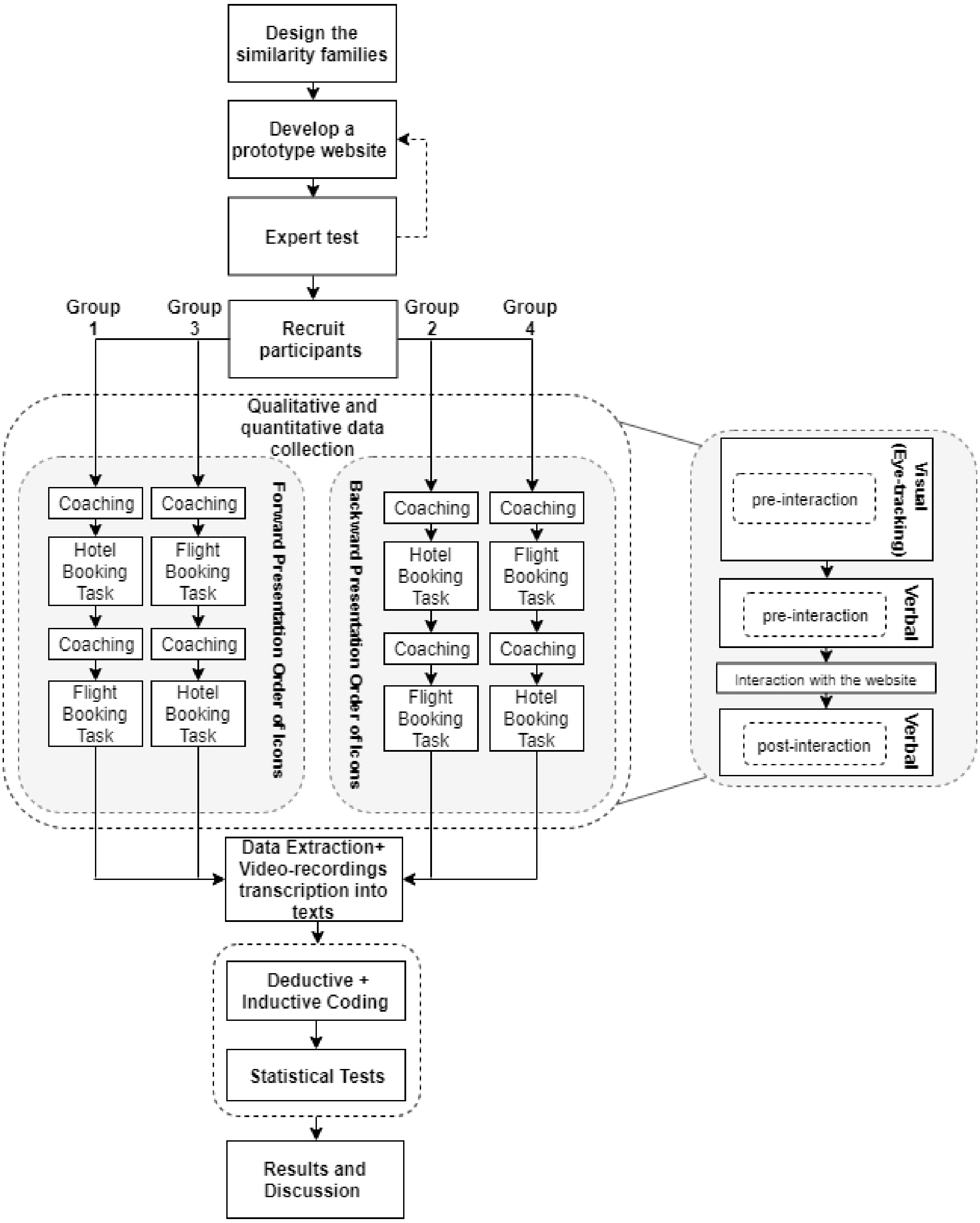


Figure 5-1: Research process of Study II.

5.2.2 Participants

A convenience sample of 24 participants was used in Study II. 22 participants were undergraduate and postgraduate students at City, University of London. Two participants were employees at City, University of London; one was an academic and the other was a non-academic member of staff. Flyers were posted on public notice boards at City, University of London. The participants were healthy adults. Participants were required to be at least 18 years old and have normal or corrected to normal vision. If someone wore eyeglasses or contact lenses, they were eligible to take part in the study. However, if they had a vision problem such as colour deficiency (colour blindness), a lazy eye, if they wore glasses with more than one power (i.e. bifocals and trifocals), or if they had eye surgery (i.e. cataract or LASIK) they could not participate. Participants received a cash gift of £10 as thanks for taking part in Study II.

5.2.3 Materials

5.2.3.1 Icons, Expert Tests and Prototype

In Study II, icons were organized into conceptual similarity families. A *similarity family* was defined as a group of icons that shared visual and semantic characteristics. Icons were designed in a similarity family as similar icons may have different and discrete purposes.

In this section, the design of the visual and semantic similarity of the icons in Study II is reported. Then, expert tests that were conducted to assess the designed icons from visual and semantic similarity perspectives are reported. Finally, the prototype website that was designed for participants to undertake the tasks in this study is presented.

5.2.3.1.1 Visual Similarity of Icons

Study II used a different set of icons to Study I. Study II had two tasks types: flight and hotel. Each of the two tasks types had three icon families. The hotel task had families for review, like, and map. The flight task had families for the user account, email, and password. Therefore, there were six icon families in total (two tasks x three families = six icon families). Each of the six families consisted of a ‘target’ and three icons which varied in levels of visual similarity to the ‘target’ icon.

Each family consisted of four icons that had varying degrees of visual similarity to each other. Visual similarity is equated to the number of characteristics shared by two objects (Taylor, 1995). For this thesis, a unique method of determining visual similarity was created. In this method, the number of visual fractions that were shared between two icons was used to determine the level of visual similarity between the two icons. A visual fraction is defined as a single or compound shape that represents part of the whole icon’s shape. The visual similarity of icons is designed according to the following similarity criteria:

1. Form 1 (slightly different): one fraction of the icon is different from a ‘target’ icon.

2. Form 2 (mostly different): two or more fractions of the icon are different from a 'target' icon. It shares one or more fractions with the 'target' icon.
3. Form 3 (completely different/dissimilar): the icon does not share any fraction with the 'target' icon. It is a totally different icon. It does not share any fraction with the 'target' icon.

Therefore, if two icons only differ by one visual fraction they are considered as slightly different visually. If two icons differ by two visual fractions but share one or more fractions, they are considered mostly different visually; and finally, if two icons share no visual fractions, they are considered completely different/dissimilar. This is illustrated in Figure 5-2 where the different fractions are circled in red. In Figure 5-2, for the review icons (icons on the left) the 'target' icon and its related icon differ by one visual fraction (the P inside the magnifying glass), hence this icon is considered slightly different from the 'target' icon. For the e-mail icons (the icons in the middle), the 'target' icon and its related icon differ by two visual fractions (single shape: envelope edge type, compound shape: up and down arrow lines inside the envelope head), hence this icon is considered mostly different from the 'target' icon. For the user account icons (icons on the right side), the 'target' icon and its related icon share no visual fractions, hence this icon is considered dissimilar from the 'target' icon.

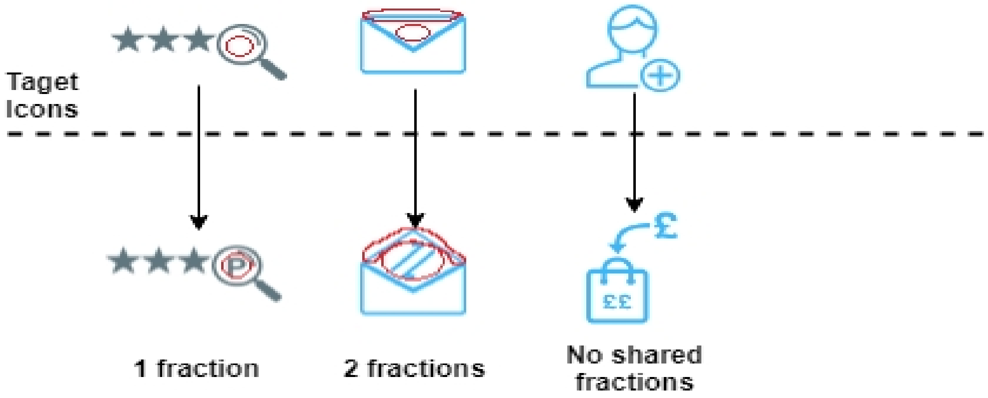


Figure 5-2: Example of icons showing how visual fractions were used to determine the visual similarity between 'target' icons and the icons related to the 'target' icons.

In each family, each icon was designed on a specific icon type (one identical icon, one slightly different icon, one mostly different icon, and one dissimilar icon). Six icon families were designed to serve in a specific presentation order. Study II had two presentation orders. However, only icons of types slightly different and dissimilar were visually different between the two presentation orders; whereas icons of types identical and mostly different were the same in both presentation orders. Therefore, six icon families were designed to serve in the 'forward' presentation order which resulted in the design of 24 icons. For the ‘backward’ presentation order, there were six families of icons, but only two icons for each family were additionally designed and the other two icons were the same as the icons designed for the ‘forward’ presentation order which resulted in the design of 12 icons. In total, 36 icons were designed to serve for both presentation orders (see Table 5-1 and Appendix-5.1 (a and b) for more details about which icons were different and which icons were the same in both presentation orders).

5.2.3.1.2 Expert Testing for Visual Similarity

The designed icons in the previous section were chosen for the experiment in Study II after expert tests. Expert tests were conducted to validate the design of the visual similarity between icons and the ‘target’ icon using fractions. Systematic questionnaires covering all 12 families of icons (six families for each presentation order) were developed and distributed to two experts. In the questionnaires, the experts were provided with the images of a ‘target’ icon and three other similar icons. Then, they were asked to rate the visual similarity of the icons to the ‘target’ icon using the visual fraction definitions given in the visual similarity of icons (section 5.2.3.1.1). To reiterate, these options were:

1. An icon is slightly different to the ‘target’ icon, if one fraction of the ‘target’ icon is different.
2. An icon is mostly different to the ‘target’ icon, if two or more fractions of the ‘target’ icon are different. It shares one or more fractions with the ‘target’ icon.
3. An icon is completely different (dissimilar) to the ‘target’ icon, if it is a completely different icon comparing to the ‘target’ icon. It does not share any fraction with the ‘target’ icon.

For an example of an evaluation that was given during the expert test for visual similarity for the icons used in the ‘forward’ presentation order, see Appendix-5.3. For the ‘backward’ presentation order, see Appendix-5.4.

The results of this test indicate that the icons were successfully designed according to the similarity criteria in this research (for the results of the expert tests see Appendix-5.3(a and b) and Appendix-5.4 (a and b)). Also, using fractions to design similar and dissimilar icons to a ‘target’ icon was successfully validated by the experts. The results showed that both experts’ responses for all icons families matched the relevant similarity and dissimilarity levels in which the icon families were designed.

After assessing the visual similarity between icons by the experts, all the assessed icons were used in the experiment of Study II without any change.

5.2.3.1.3 Semantic Similarity of Icons

The meaning of an icon was designed according to the visual representation of that icon. Semantic similarity between icons might affect the outcomes of this research. As seen in Study I, two visually similar icons that had two different/dissimilar meanings misled/confused the users. Therefore, the semantic similarity of icons in each family was controlled by designing the icons to be semantically like a certain degree. The reason for controlling the semantic similarity between icons was to focus on the relationship between the IVs and the DVs. To control the semantic similarity between the meanings of the icons in the families, the icons were assigned semantic similarity values (SSVs) based on the text descriptions of the meanings of the icons. SSVs were calculated for the textual descriptions for the following pairs in each family:

1. ‘Target’ and slightly different icons.
2. ‘Target’ and mostly different icons.
3. ‘Target’ and dissimilar icons.

This was done by using the greedy pairing method with WordNet[5] (which was discussed in Chapter 3) (Lintean and Rus, 2012). WordNet was developed at Princeton University and is an English lexical reference system. WordNet can be used to compute SSVs between two texts by comparing the similarity between a pair of words (Bogdani, 2016; Run, 2014), with higher scores indicating greater similarity. The work for this thesis utilised the greedy pairing (or greedy matching) algorithm (see Chapter 3) when calculating the SSVs. For this thesis, the textual definition of ‘target’ icons was compared to the textual definition of icons which were visually similar to the ‘target’ icons. Only icons with SSVs more than or equal to 0.5 were placed in the same family. Within a family, the icons are visually coherent, but individual icons have different, discrete purposes. To see the visual representations, meanings of the icons, and the calculated SSVs, refer to Appendix-5.1 (a and b).

To ensure that the icons that belonged to different families were semantically dissimilar (SSVs less than 0.5), randomly selected icons were chosen and SSVs of those icons were calculated (see Appendix-5.2). The results indicated that two randomly selected icons from two different families had lower SSVs than 0.5.

---

[5] This can be accessed online: http://deeptutor2.memphis.edu/Semilar-Web/public/demo.jsp.

5.2.3.1.4 Expert Testing for the Textual Descriptions of the Icons' Meanings

Expert tests were also conducted to ensure that the textual descriptions given to the icons matched the actual function/behaviour of the icons on the website. Systematic questionnaires covering all six families of icons were developed and distributed among the two experts along with the developed prototype. In the questionnaires, the experts were provided with the images of the icons and textual descriptions for the functions of the icons. Then, they were asked to go through the website and, in parallel, check the actual functions of the icons in the website (when they click on the icons) with the textual descriptions provided in the questionnaire. They rated the textual descriptions for the functions of the icons using the following options:

1. Correct**:** the textual description of the function of an icon is *correct* and the function of the icon was *correctly* designed in the user interface.
2. Incorrect description: the textual description of the function of an icon is partly or completely incorrect. Thus, the text must be re-described.
3. Change/improve the design of user interface (UI): the function of an icon is *correct*ly described in the text, but the design of the user interface needs some improvements (i.e. the email address must be clearly specified in the sending form).

For an example of an evaluation given during the expert test for the textual description of the meanings of the icons that were used in the experiment, see Appendix-5.5 (a and b).

The results of this test indicate that the textual descriptions of the meanings of the designed icons were correct and clearly described what those icons do on the website. No incorrect description was reported by the experts. However, the experts suggested improving the design of the user interface. The suggestions were related to the colours or the size of the fonts used in the website. One of the experts suggested the following:

(1) 'Write subject in full'.
(2) 'Change the font colour in the description'.
(1) 'The font size in the description is very large'.

All the suggestions were considered to improve the design of the website.

5.2.3.1.5 Prototype Website and Orders

A new prototype website for 'holiday booking' was created for Study II. The experiment was conducted on a PC. However, as this research is being investigated in the context of limited-size screens, the prototype website was displayed on the PC screen using a mobile interface emulator (*responsinator* as an extension to the Chrome web browser). An iPhone 7 interface option with 375x667 user interface kit size (points) was used. The remainder of the PC screen was black to allow participants to focus on the displayed mobile interface. Figure 5-3 and Figure 5-4 show the main web-pages for the flight and the hotel booking tasks in the 'forward' presentation order.

Participants were first presented with a known 'target' icon and then, depending on the viewing order condition they were in, they saw an icon which was either slightly different or dissimilar to the known 'target' icon shortly after viewing the known 'target' icon. Participants' speed and accuracy of identifying the similar or very different icons were recorded. As discussed in the study design (section 5.2.1), it was necessary to counterbalance the order in which icons were presented, and two presentation orders ('forward' and 'backward') were used.

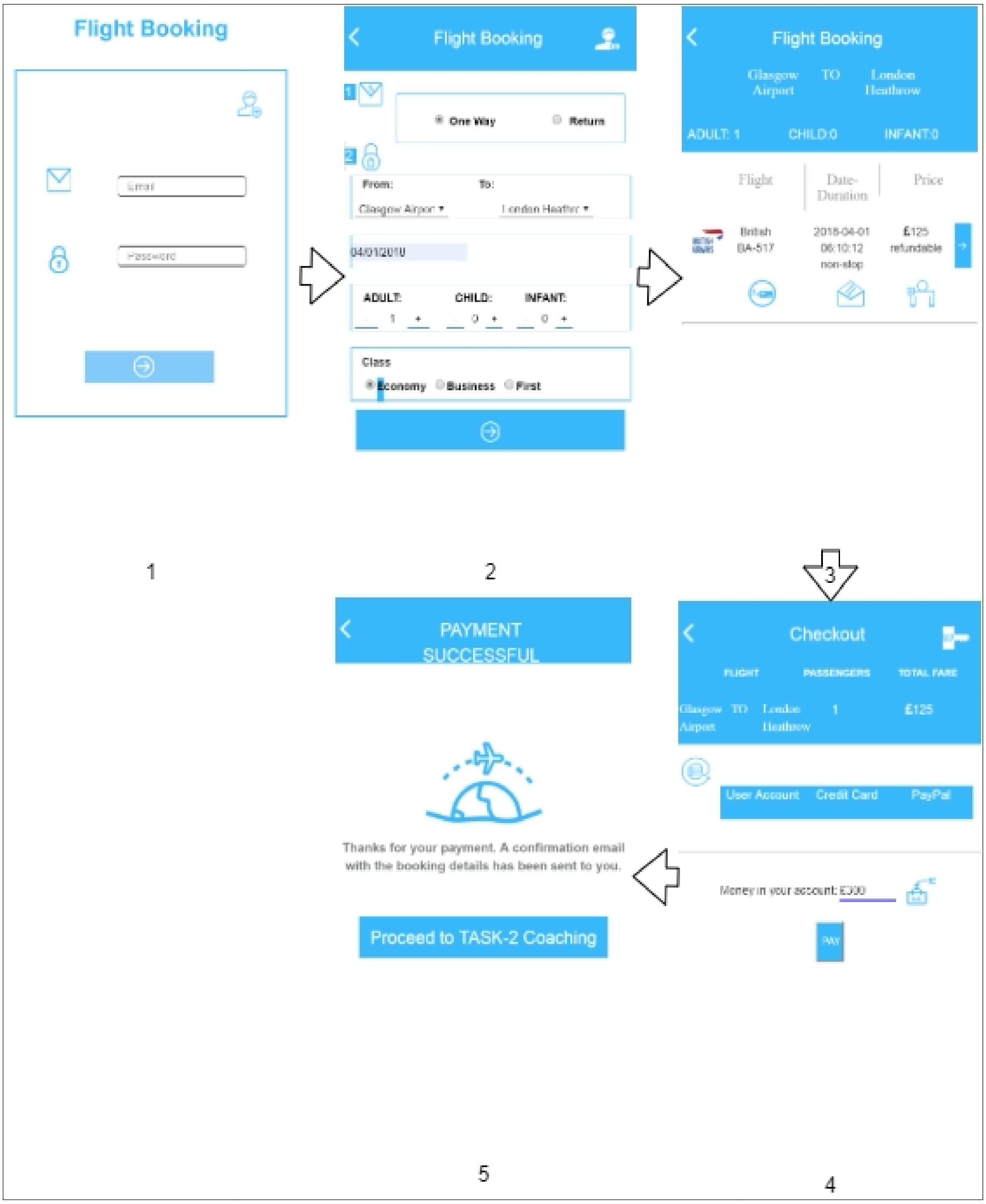
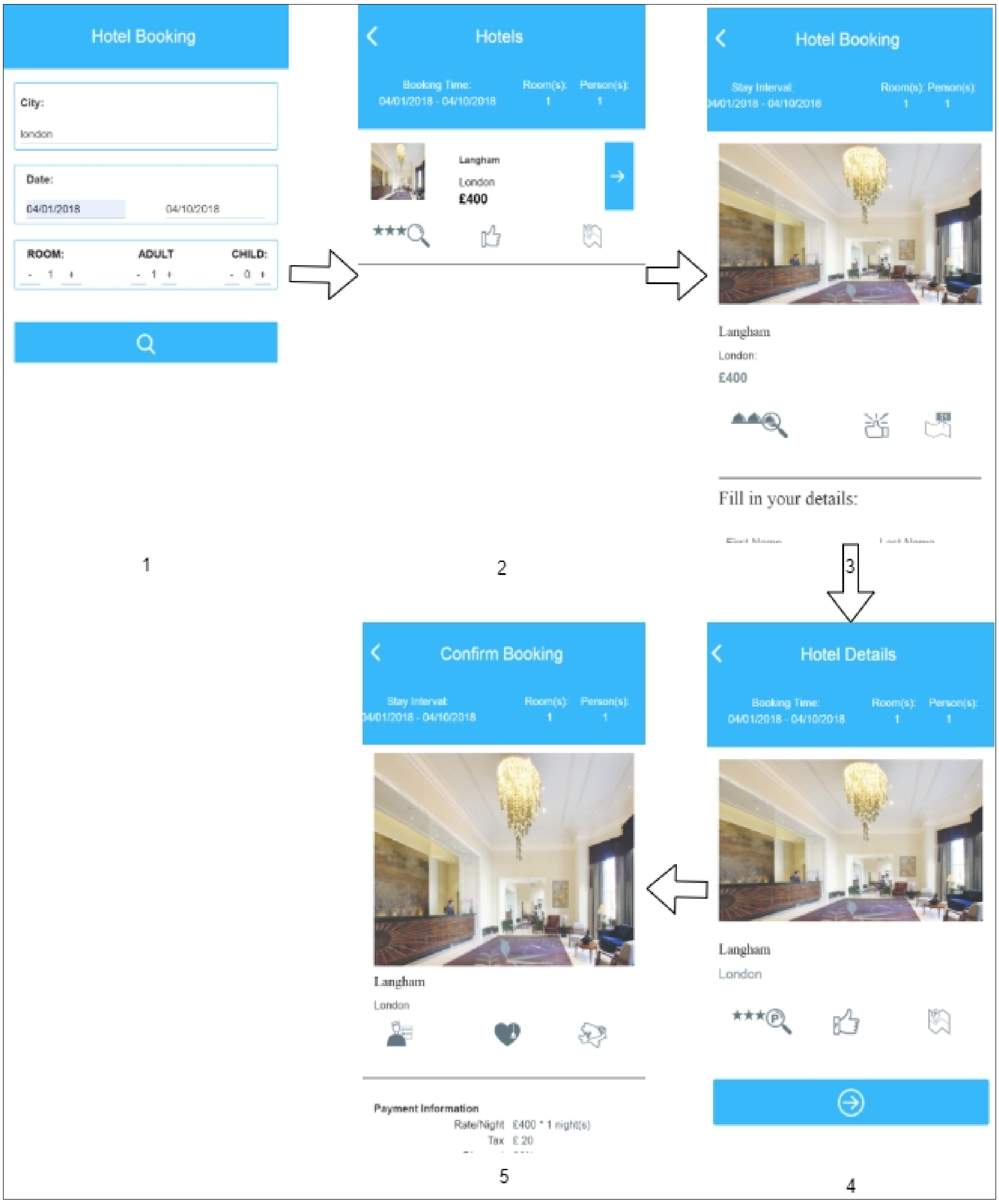


Figure 5-3: Screenshots of the main web pages in the flight booking task (‘forward’ presentation order).



City: london

Date: 04/01/2018 04/10/2018

ROOM: - 1 + ADULT - 1 + CHILD: - 0 +

Booking Time: 04/01/2018 - 04/10/2018 Room(s): 1 Person(s): 1

Langham London £400

Stay Interval 04/01/2018 - 04/10/2018 Room(s): 1 Person(s): 1

Langham London: £400

Fill in your details:

Booking Time: 04/01/2018 - 04/10/2018 Room(s): 1 Person(s): 1

Langham London

Stay Interval: 04/01/2018 - 04/10/2018 Room(s): 1 Person(s): 1

Langham London

Payment Information

Rate/Night £400 * 1 night(s)

Tax £ 20

Figure 5-4: Screenshots of the main web pages in the hotel booking task (‘forward’ presentation order).

For the ‘forward’ presentation (see Figure 5-5), icons were presented in the following order: identical, slightly different, mostly different, and dissimilar. This order presented the participant with a gradually diverging pattern of icons. In the ‘backward’ presentation (see Figure 5-6), the icons were presented in the following order: identical, dissimilar, mostly dissimilar, and slightly different. This order presented the participant with a gradually converging pattern of icons.

The ‘target’ icons were icons which are commonly used on the internet on websites and mobile applications, whereas the other forms were designed according to the similarity criteria mentioned earlier. The meanings of the slightly different, mostly different, and dissimilar icons were designed according to the visual representations of those icons and inspired by the meanings of the ‘target’ icons. The 24 participants were divided evenly across the four groups (see Table 5-1), resulting in six participants per group. The total number of 24 participants is enough in many cases for statistical testing. Previous studies have considered different sample sizes. Freeman *et al.* (2009) conducted a comparative study on 22 participants evenly divided into two groups (11 participants per group). In another study, De Angeli *et al.* (2006) conducted a usability evaluation study on 28 participants based on two experimental conditions (menu-based and metaphor-based).

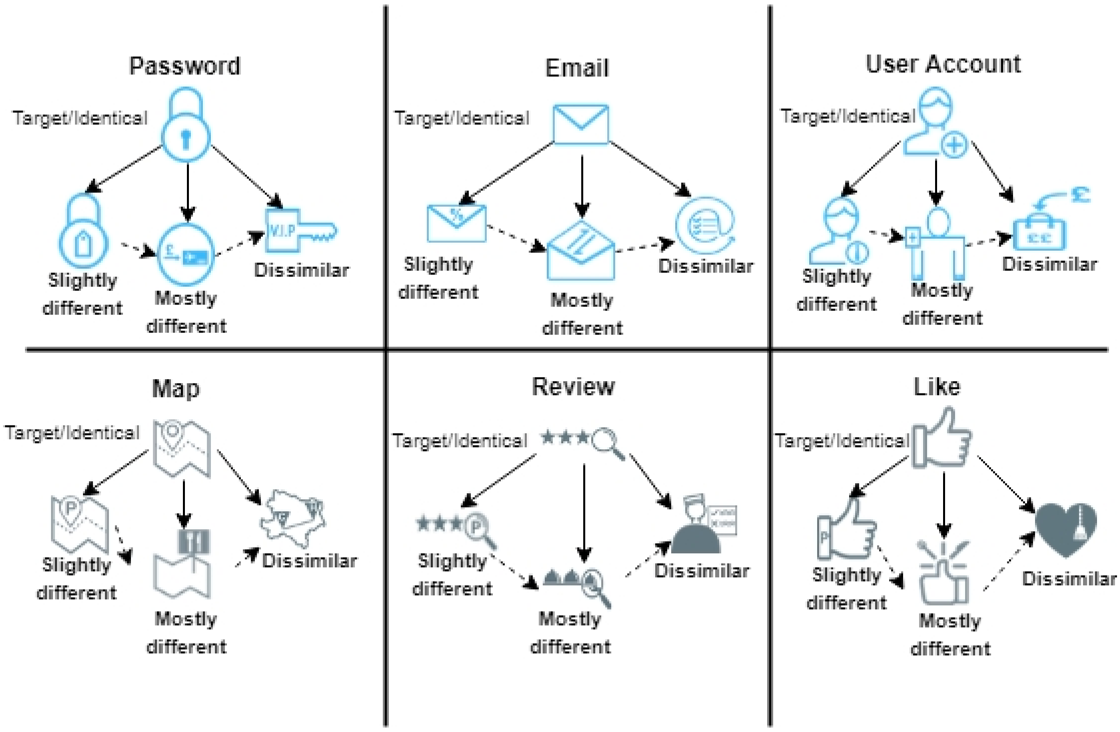


Figure 5-5: The similarity families used in the ‘forward’ presentation order. There were six families of icons (password, email, user account, review, map, and like). Each family has a ‘target’/identical icon and three variations. The identical icons are visually identical to the ‘target’ icon. The slightly different icons are visually slightly different to the ‘target’ icon. The mainly different icons are visually mostly different to the ‘target’ icon. The dissimilar icons are visually dissimilar to the ‘target’ icon.

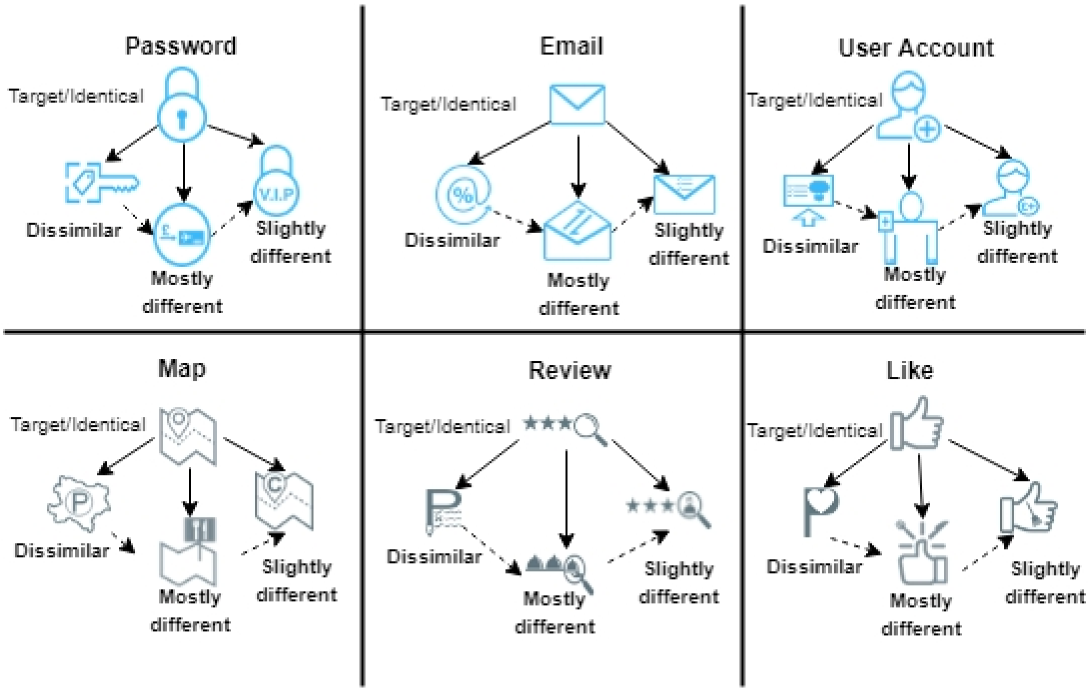


Figure 5-6: The similarity families used in the ‘backward’ presentation order. There were six families of icons (password, email, user account, review, map, and like). Each family has a ‘target’/identical icon and three variations. The identical icons are visually identical to the ‘target’ icon. The slightly different icons are visually slightly different to the ‘target’ icon. The mainly different icons are visually mostly different to the ‘target’ icon. The dissimilar icons are visually dissimilar to the ‘target’ icon.

5.2.4 Visual and Verbal Data Collection: How To?

In Study II, qualitative and quantitative data were collected. Qualitative data were collected in the form of participants' verbalisations while they gave a *think-aloud* both before and after they interacted with icons. These data were collected to determine the participants’ comprehension of the icons. Quantitative data were collected using an eye-tracker (visual data). Eye-tracking was used to observe participants’ eye-movements during the study to determine how long participants looked at an icon. However, if verbal and visual data are collected simultaneously, there can be interference between the data types. To avoid this interference, it is common in saliency research studies for researchers to use free-viewing protocols. Some researchers use a fixed-time interval free-viewing protocol, while others use an open-ended free-viewing protocol.

In a fixed-time interval protocol, the participant is allowed to view the stimuli for a fixed period that the researcher has pre-determined (for example, three seconds, five seconds, one minute). A participant may view a computer screen with many objects and be asked to find a red square. In the fixed-time interval method, the participant can view the screen for a set amount of time (e.g., three seconds). Judd *et al.* (2012) used a free-viewing protocol conducted at a three-second fixed time interval. Borji and Itti (2015) used a free-viewing protocol conducted at a five-second fixed time interval.

In the open-ended time protocol, the participant can view the stimuli until they decide they no longer need to view the stimuli, which often results in the participant deciding about the stimuli. For example, a participant may view a computer screen with many objects and be asked to find a red square. In the open-ended viewing method, the participant can view the screen until they have found the red square or decide that they cannot find it. Ehinger *et al.* (2009) allowed their participants to view stimuli until the participant pressed a key on the keyboard which terminated the viewing phase. Koehler *et al.* (2014) let their participants view stimuli until they (the participants) decided whether a cue object was present or absent.

Study II utilised a two-phase data-collection process that allowed the researcher to collect quantitative eye-tracking data in the viewing phase and qualitative data about the participants’ understanding of the meaning of the icons, along with their reasons for interpreting the icons the way they did, in the think-aloud phase. This design was used to avoid interference between the users’ visual and verbal data during the visual data collection phase. Furthermore, by allowing the participants to take their time to view the icons, individual user patterns were shown. In this research, accuracy was more important than speed, therefore, an open-ended viewing method was used so that the participants would not feel stress or pressure to complete the task.

5.2.5 Apparatus

A Tobii X-60 eye-tracker was used to collect data about the participants' eye movements. It was utilised to collect the total eye fixation durations (viewing times) on icons by defining them as areas of interest (AOIs). The eye-tracker software used was Tobii Studio 3.3.2. Eye-tracking provided additional information for this study, because it allowed the researcher to determine where exactly the participants looked at an interface and to quantitatively measure how long the participants viewed the various icons.

5.2.6 Procedure

Information about the experiment and the tests that would be involved were sent to participants via email before the experiment. This included the participant information sheet and the guidance sheet. Different guidance sheets were sent based on the group the participant would be in (hotel-first in Appendix-5.6 and flight-first in Appendix-5.7). Study II took place at City, University of London. Participants signed a consent form before starting the experiment. During the experiment, the participants were told to assume they were alone to avoid the distraction of the researcher's presence. The experiment was 30 to 40 minutes long and consisted of five main phases: 1. calibration of the eye-tracker; 2. coaching for the task 1; 3. performance of task 1; 4. coaching for task 2; 5. performance of task 2 (note the order of task presentation depended on the group the participant was in).

Before the experiment, the participants were given instructions about what to do and what not do during calibration of the eye-tracker (Appendix-5.8), then the calibration was performed. Before beginning their coaching on the first task, participants were given the following instructions on the screen:

- You need to follow this protocol on each screen (the three-step protocol):

1. There are icons on the following screens. Look at them in ***silence*** and think of what these icons are going to do if you click on them. Kindly say ***'I am done'*** when you are finished looking.

2. Tell me:
    a. What do you think will happen if you clicked on each icon?
    b. Why do you think that will happen?

3. Click on each icon then,
    a. Tell me what happened?

Step one of the protocols consisted of open-ended viewing of the icons (in silence). In this phase, participants looked at the icons on the screen in silence and thought about the meanings of these icons. Participants could view the icons for as long as they wanted and were instructed to verbally state, ‘I am done’, when they were done viewing the icons. Steps two and three of the protocol, the think-aloud steps, consisted of the participants saying aloud what they thought the meanings of the icons were.

After reading the instructions, the participants clicked a button at the bottom which said, ‘START COACHING’, in order to start their coaching. The icon that was seen in the coaching phase was shown again at the beginning of the task phase. During the coaching for the flight task, the participants saw a screen that said, ‘You need to login’. They were given an email id and password to use. Then they clicked a button at the bottom of the screen which said, ‘START’. They then practised the flight booking task where they engaged in the three-stage protocol (the three-step protocol) mentioned in the previous paragraph. During coaching for the hotel task, the participants saw a screen that said, ‘You will book the following hotel’. They were given the name of a city, check-in and check-out dates, the number of rooms to be booked, and the number of people staying. Participants then clicked a button at the bottom of the screen which said, ‘START’. They then practised the hotel booking task where they engaged in the three-step protocol. At the end of either coaching session, the participants were presented with a screen that said, ‘You have completed the coaching session’. Then they clicked a button at the bottom of the screen which said, ‘START TASK’ (the button said TASK 1 or TASK 2, depending on the task the participant had just completed). After coaching, the participants performed the tasks. The verbal responses of the participants given during the experiment were audio and video recorded.

5.2.6.1 Initialisation Process

Participants may or may not have experienced the ‘target’ icons before. Prior experience with icons is an influential factor in the performance of participants (Zhang and Ghorbani, 2004). In experimental tests, data must be obtained in equivalent conditions and the impact of other factors must be eliminated. Thus, an initialisation process was conducted by repeating the exposure of ‘target’ icons (the first time was in the coaching process and the second time was at the beginning of the task) to achieve the following:

(1) Assure equal experimental conditions for all participants; thus, icons on identical forms are experimentally learned by all participants.

(2) Utilise the participants' viewing times on identical icons as the base values to measure the speed with which the participants recognise visually similar icons.

At the initialisation phase, participants were expected to recognise the icons efficiently and fast with a high accuracy rate of interpretation. In Study II, the initialisation process was established by repeating the exposure of the icons during the coaching process and again during the relevant task. Table 5-2 shows the coaching process before each task for all groups.

Group	**Presentation order**	**Coaching**	**Task1 (Block)**	**Coaching**	**Task2 (Block)**
1	Forward	Target	Hotel	Target	Flight
2	Backward	Target	Hotel	Target	Flight
3	Forward	Target	Flight	Target	Hotel
4	Backward	Target	Flight	Target	Hotel

Table 5-2: Shows how the participants were coached before conducting the hotel and flight task types in the four groups.

5.2.7 Data Pre-Processing

5.2.7.1 Qualitative Data Coding

The procedure that was used in Study I for coding the qualitative data was also used in Study II. The collected qualitative data were coded to create quantitative data for analysis. The participants' accuracy of comprehension was scored on a five-point accuracy scale. The four groups were compared on their levels of accuracy of interpretation of the icons.

The responses for question 1, 'What do you think will happen if you clicked on each icon?', were deductively coded on the same five-point scale as was used in Study I. This scale was: accurate = 5, moderate = 4, conflicting = 3, erroneous = 2, and incapable = 1. The responses for question 2, 'Why do you think that will happen?' were inductively coded (Islam and Tetard, 2014); this process was the same as Study I. For question 3, the participants were first asked to click on the icon and then were asked to tell the experimenter what happened. For question 3, 'What do you think happened when you clicked on the icon?' in case the participants weren't accurate, they were told the accurate answer. No coding was done for question 3 as most participants were accurate at learning what happened during post-interactions.

The accuracy of interpretation of the user account, email, password, review, like, and map were averaged to get a mean accuracy value for each icon type for each participant. All the accuracy of interpretation values was tested for their normality.

5.2.7.2 Quantitative Data Extraction - Viewing Times

The eye fixation durations of the participants were extracted from the eye-tracker. The duration of the viewing in silence phases was determined in Tobii Studio 3.3.2 by logging timestamps. Timestamps were determined based on two events. The *LeftMouseClick* event was enabled by Tobii Studio. The ‘I am done’ event was manually logged when the participants said, *‘I am done’*. The metric used was *fixation duration*.

Viewing in silence phases were tagged for each screen. For a screen, the viewing in silence phase start time is tagged when the *LeftMouseClick* event (the left mouse click that moved a participant from a previous screen to a current screen) happened. This event precedes loading a current webpage. The end time of the viewing in silence phase is tagged when the participant says, *‘I am done’* (see Figure 5-7). After the experiment, the participants’ viewing times on the screen were extracted and the logged timestamps converted into seconds.

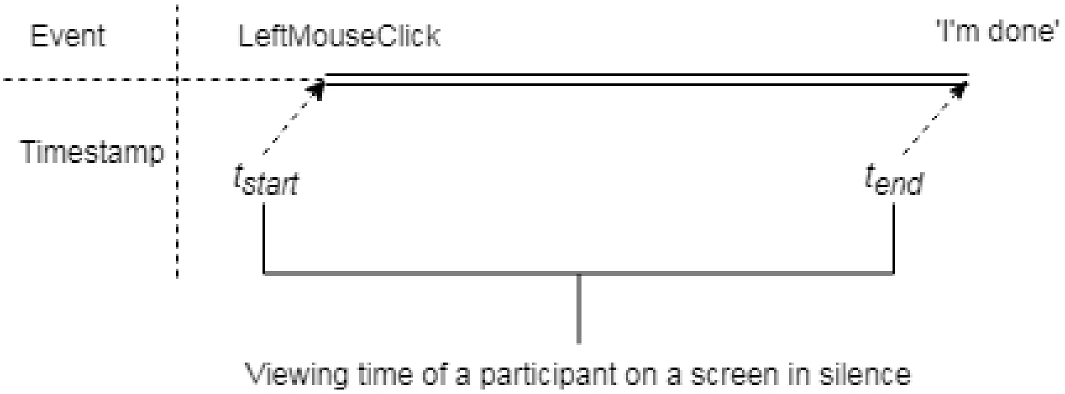


Figure 5-7: Time-stamping viewing in silence phase.

To collect participants’ viewing times on icons across all screens, Static AOIs were drawn on the icons. Depending on the shape of the icons, the suitable shapes of AOIs were chosen. The eclipse and the rectangle shapes of AOIs were used and they were carefully positioned on the icons to ensure that these shapes covered the shape of the icons and to allow precise collection for viewing times.

AOIs were replicated using *copy* and *paste* commands when the screens were different and *copy* and *paste_in _place* if the screens were the same. (i.e. coaching screen and the first screen of a task are same). This allows the precise collection of viewing times on icons across all screens. When an AOI is drawn, it is made activate during the viewing time slot on a screen. After that, fixation duration is selected in the metric menu. In descriptive statistics, N (the number of fixations on an AOI), max (the maximum length of fixations on an AOI), min (the minimum length of fixations on an AOI), and SUM (the sum of fixations durations on an AOI) are selected. Using media time, t_{start}, denotes the *start time* of the media (in this research, it is the start time of the viewing on a screen in silence); and *duration* is the difference in time between t_{end} and t_{start}in seconds.

The participants' fixation durations on icons/AOIs were extracted from the eye-tracker. These data were tabulated for each individual participant and for each icon. The fixation duration data were then processed to calculate the mean fixation duration times for both the flight icon and hotel icon families. The viewing times of user account, email, password, review, like, and map were averaged to get a mean viewing time value for each icon type for each participant. All the viewing time values were tested for their normality.

5.2.8 Data Analysis

The collected data were extracted from the eye-tracker and the audio-recordings were transcribed into texts. Afterwards, the transcribed data were coded using the five-point scale that was discussed earlier. In this section, the normality of the data and choosing the right statistical tests are discussed. For each research question, the hypothesis and the null hypothesis are investigated. The requirements for the statistical tests are also discussed.

For all the research questions, to prevent any confounding effects (i.e. learning effects), the data from the two task types (hotel and flight) were combined in the analysis. Moreover, for research questions RQ2.1, RQ2.3 and RQ2.5, to prevent any confounding effects (i.e. recency effects), the data of the two presentation order groups ('forward' and 'backward') was combined in the analysis.

The IVs were:

(1) *Presentation order*, which had two levels: 'forward' and 'backward', and

(2) *Visual similarity/icon type*, which had four levels: ‘identical’, ‘slightly different’, ‘mostly different’, and ‘dissimilar’.

The DVs were:

(1) *Speed of recognition*, which was reflected by the participants’ viewing time of the icons measured in seconds (investigated in RQ2.1, RQ2.2, and RQ2.5).

(2) *Accuracy of interpretation* (investigated in RQ2.3, RQ2.4, and RQ2.5).

5.2.8.1 RQ2.1: Will participants who are presented with visually similar icons in a ‘forward’ order recognise the icons at a different speed than participants who are presented with visually similar icons in a ‘backward’ order?

RQ2.1 was addressed by investigating the effects of icon presentation order and visual similarity of icons on participants’ speed of recognition of the icons. The hypothesis of RQ2.1 will focus on presentation order; while RQ2.2 will focus on icon type and present a hypothesis thereon. Furthermore, the Latin square design of this study did not assume an interaction between IVs and confounds all ‘main’ effects with the ‘interaction’ effects among two or more IVs (Kohli, 1988). To answer RQ2.1 and RQ2.3 which involves two IVs, two-factor (factorial) analysis is conducted to show ‘main’ and ‘interaction’ effects of IVs when measuring the DVs (the speed of recognition in RQ2.1 and accuracy of interpretation in RQ2.3). ‘Main’ effect is defined as the effect of one of the IVs on the DV, regardless of the effect of any other IV. ‘Interaction’ effect is defined as the change of the effect of one IV on the DV at a different level to another IV. In this research, no hypothesis about ‘interaction’ effects were made; but rather arrangements in the analysis to show realistic ‘main’ effects of presentation orders on the DVs. The hypothesis and the null hypothesis of RQ2.1 were:

- **$H_{2.1}$:** Participants who view the icons in the ‘forward’ presentation order will be significantly faster at recognising the icons than participants who view the icons in the ‘backward’ presentation order.
- **H_0:** The ‘forward’ and ‘backward’ presentation order groups will not recognise the icons at a significantly different speed.

To address the hypothesis of RQ2.1, the data from the two task types (hotel and flight) was combined. The analysis for RQ2.1 focused on comparing the participants’ speed of recognition for the icons across the two presentation orders (‘forward’ and ‘backward’). However, since icon type may have influenced the participants’ speed of recognition, icon type was also included in the analysis. Icon type will be investigated further in RQ2.2. The initial analysis plan for RQ2.1 was to conduct a mixed ANOVA. A mixed ANOVA compares the mean differences between groups that have been split into two factors, where one factor is a within-subjects factor (here, icon type) and the other factor is a between-subjects factor (here, presentation order).

Before conducting a mixed ANOVA, the assumptions for this ANOVA type were verified:

1. One assumption of a mixed ANOVA is that there are no significant outliers in any group of the within-subjects factor or the between-subjects factor. To test for this, the data points were graphed in boxplots and z-scores were calculated. The boxplot (Appendix-5.9) and z-scores for the ‘backward’ presentation order, identical icon type indicated that there was one significant outlier (z-score +/- 2.68) and one nearly significant outlier (z-score = 2.23). Since there were only 24 participants in the study, and 12 participants per presentation order, it was decided that removing the outliers was not a good option as it would reduce the sample size.
2. The next assumption for a mixed ANOVA is a normal distribution of the dependent variable for each combination of the groups of the two factors (within-subjects and between-subjects). To check if the data were normally distributed, a Shapiro-Wilk normality test was run (this is the preferred test when there are less than 50 participants). The Shapiro-Wilk test (see Appendix Appendix-5.10) and histogram (see Appendix-5.11) showed that for the ‘backward’ presentation order, identical icon type the data were not normally distributed, which was likely due to the two outliers.
3. Another assumption for the mixed ANOVA is the homogeneity of variance for each combination of the groups of the two factors (within-subjects and between-subjects). To test this assumption, Levene’s test for homogeneity of variances was conducted. It was found that there were no violations of homogeneity of variance.
4. The final assumption for a mixed ANOVA is sphericity, that the variances of the differences between the related groups of the within-subject factor for all groups of the between-subjects factor must be equal. To test this assumption, Mauchly's test of sphericity was conducted. For the ‘forward’ presentation order, Mauchly's test of sphericity was not significant; however, for the

> ‘backward’ presentation order, it was, indicating that the assumption of sphericity had been violated (χ2(5) = 16.66, p = .005).

After testing the assumptions for a mixed ANOVA, it was found that there were significant outliers in the ‘backward’ presentation’s identical icon group. Furthermore, it was found that the DV was not normally distributed for the ‘backward’ presentation’s identical icon group (likely due to the outliers). Furthermore, the assumption of homogeneity of variance was violated by the ‘backward’ presentation order. Although these violations probably resulted from the two outliers, the researcher chose not to remove them due to the already small sample size. Therefore, a non-parametric test was needed. Possible non-parametric alternatives to a mixed ANOVA include the Scheirer-Ray-Hare Test (Zaiontz, 2012) and Aligned Rank Transform tool (ARTool) (Wobbrock *et al.*, 2011). The ARTool was designed as a general non-parametric factorial analysis to be used in human-computer interaction (HCI) studies; therefore, this method was used to analyse the data for RQ2.1.

According to the creators of the ARTool (Wobbrock *et al.*, 2011), a popular method for dealing with non-parametric data is to use Conover and Iman’s (1981) Rank Transform (RT) method. This method consists of applying ranks, which are averaged in the case of ties, over an entire data set, then applying a parametric ANOVA to the ranked data. This results in a non-parametric factorial procedure. This procedure is reliable for ‘main’ effects, but not for interactions; interactions that are assessed in this manner are subject to Type I errors (Salter and Fawcett, 1993; Higgins and Tashtoush, 1994). The ART procedure was devised to correct this problem.

Wobbrock *et al.* (2011) describe how the ARTool works:

> For each ‘main’ effect or ‘interaction’, all responses (Y_i) are ‘aligned’, a process that strips from Y all effects but the one for which alignment is being done (‘main’ effect or ‘interaction’ effect). This aligned response we’ll call $Y_{aligned}$. The aligned responses are then assigned ranks, averaged in the case of ties, and the new response we’ll call Y_{art}. Then a factorial ANOVA is run on the Y_{art} responses, but only the effect for which Y was aligned is examined in the ANOVA table. Thus, for each possible ‘main’ or ‘interaction’ effect, one new aligned column ($Y_{aligned}$) and one new ranked column (Y_{art}) is necessary. For example, with two factors and their interaction, we need six additional columns: three aligned and three ranked, where each set of three comprise each of two factors and their interaction. In general, for N factors, we need 2N-1 aligned

columns and 2N-1 ranked columns. Because creating these columns is tedious, the program provided here, ARTool, creates these columns for you. This is the main function of ARTool.

An example of ARTool's output for the speed of recognition data is given in Appendix-5.12.

5.2.8.2 RQ2.2: What is the impact of the degree of visual similarity between icons on the speed with which a participant recognises the icons?

RQ2.2 was addressed by investigating the effects of icon visual similarity on participants' speed of recognition of four icon types. The hypothesis and the null hypothesis of RQ2.2 were:

- **$H_{2.2}$:** The participants' speed of recognition will be significantly different between the four icon types, in which participants will recognise the icons that are identical to the known 'target' icon the quickest, participants will recognise the slightly different icons slower than the identical icons, participants will recognise the mostly different icons slower than the slightly different icons, and finally, participants will recognise the icons that are dissimilar to the known 'target' icon the slowest.
- **H_0:** The participants' speed of recognition will not be significantly different between the four icon types.

To address the hypothesis of RQ2.2, the data of the two presentation order groups ('forward' and 'backward') was combined. Furthermore, the data from the two task types (hotel and flight) was combined. The analysis for RQ2.2 focused on comparing the participants' speed of recognition across the four icon types (identical, slightly different, mostly different, and dissimilar).

The initial analysis plan for RQ2.2 was to conduct repeated measures ANOVA, due to the within-subjects design of the experiment. The assumptions for this ANOVA type were verified:

1. One of the assumptions of the repeated measures ANOVA is that there are no significant outliers. To test for this, the data points were graphed in boxplots and z-scores were calculated. The boxplots and z-scores indicated that for the identical icon type there was one significant outlier (z-score +/- 2.68) and one nearly significant outlier (z-score 2.23). The boxplot for these outliers can be seen in Appendix-5.13. Since there were only 24 participants in the study, it

was decided that removing the outliers was not a good option as it would reduce the sample size.

2. Another assumption of the repeated measures ANOVA is that the distribution of the dependent variable should be approximately normally distributed. To check if the data were normally distributed, a Shapiro-Wilk normality test was run (this is the preferred test when there are less than 50 participants). The Shapiro-Wilk test showed that the data for the dissimilar and mostly different icons were normally distributed; however, the data for the identical and slightly different icons were not normally distributed (see Appendix-5.14). The histogram for the identical icons can be seen in Appendix-5.15 and the histogram for the slightly different icons can be seen in Appendix-5.16.
3. A final assumption of the repeated measures ANOVA is that there is sphericity. To test this assumption, Mauchly's test of sphericity was conducted. Mauchly's test of sphericity was significant, indicating that the assumption of sphericity had been violated ($\chi2(5) = 19.64$, $p = .001$).

An alternative approach for analysing abnormally distributed data is to use a corresponding non-parametric test. The Friedman test is the non-parametric alternative to the one-way ANOVA with repeated measures; hence, a Friedman test was used to analyse the data for RQ2.2.

5.2.8.3 RQ2.3: Will participants who are presented with visually similar icons in a 'forward' order interpret the icons with different accuracy than participants who are presented with visually similar icons in a 'backward' order?

RQ2.3 was addressed by investigating the effects of icon presentation order and icon visual similarity on participants' accuracy of interpretation of the icons. The hypothesis of RQ2.3 will focus on presentation order, while RQ2.4 will focus on icon type and present a hypothesis thereon. Again, the Latin square design of this study did not assume an interaction between IVs (see section 5.2.8.1 for further discussion). The hypothesis and the null hypothesis of RQ2.3 were:

- **$H_{2.3}$:** Participants who view the icons in the 'forward' presentation order will be significantly more accurate at interpreting the meaning of the icons than participants who view the icons in the 'backward' presentation order.
- **H_0:** The 'forward' and 'backward' presentation order groups will not be significantly different at accurately interpreting the meaning of the icons.

To address the hypothesis of RQ2.3, the data from the two task types (hotel and flight) was combined. The analysis for RQ2.3 focused on comparing the participants' accuracy of interpretation of the icons across the two presentation orders ('forward' and 'backward'. However, since icon type may have influenced the participants' accuracy of interpretation, icon type was also included in the analysis. The between-subjects factor for this analysis was presentation order and the within-subject factor was icon type. Icon type is investigated further in RQ2.4. The initial analysis plan for RQ2.3 was to conduct a mixed ANOVA. A mixed ANOVA compares the mean differences between groups that have been split into two factors, where one factor is a within-subjects factor (here, icon type) and the other factor is a between-subjects factor (here, presentation order).

Before conducting a mixed ANOVA, the assumptions for this ANOVA type were verified:

1. One assumption of a mixed ANOVA is that there are no significant outliers in any group of the within-subjects factor or the between-subjects factor. To test for this, the data points were graphed in boxplots and z-scores were calculated. The boxplot (see Appendix-5.17) and z-scores for the 'forward' presentation order, identical icon type indicated that there was one significant outlier (z-score +/- 2.68) and two nearly significant outliers (z-scores = -0.63 and -1.71). The boxplot (see Appendix-5.17) and z-scores for the 'backward'

presentation order, identical icon type indicated that there was one significant outlier (z-score +/- 2.68). The boxplot (see Appendix-5.18) and z-scores for the ‘backward’ presentation order, slightly different icon type indicated that there were three nearly significant outliers (z-scores = -1.32, 2.14, and -1.32). The boxplot (see Appendix-5.19) and z-scores for the ‘backward’ presentation order, dissimilar icon type indicated that there was one nearly significant outlier (z-scores = 1.71). Since there were only 24 participants in the study, and 12 participants per presentation order, it was decided that removing the outliers was not a good option as it would reduce the sample size.

2. The next assumption for a mixed ANOVA is a normal distribution of the dependent variable for each combination of the groups of the two factors (within-subjects and between-subjects). To check if the data were normally distributed, a Shapiro-Wilk normality test was run (this is the preferred test when there are less than 50 participants). The Shapiro-Wilk test (see Appendix-5.20) showed that for the ‘forward’ and ‘backward’ presentation orders, identical icon type; the ‘forward’ presentation order, slightly different icon type; and the ‘backward’ presentation order, dissimilar icon type that the data were not normally distributed. These patterns were also present in the histograms: ‘forward’ presentation order, identical icon type (see Appendix-5.21), ‘backward’ presentation order, identical icon type (see Appendix-5.22), ‘forward’ presentation order, slightly different icon type (see Appendix-5.23), and ‘backward’ presentation order, dissimilar icon type (see Appendix-5.24). These patterns were likely due to the outliers.
3. Another assumption for the mixed ANOVA is homogeneity of variance for each combination of the groups of the two factors (within-subjects and between-subjects). To test this assumption, Levene’s test for homogeneity of variances was conducted. It was found that there were no violations of homogeneity of variance.
4. The final assumption for a mixed ANOVA is sphericity, that is the variances of the differences between the related groups of the within-subject factor for all groups of the between-subjects factor must be equal. To test this assumption, Mauchly's test of sphericity was conducted. Mauchly's test of sphericity was not significant for the ‘forward’ or ‘backward’ presentation orders, indicating that the assumption of sphericity had not been violated.

After testing the assumptions for a mixed ANOVA, it was found that there were significant outliers in the 'forward' and 'backward' presentation order identical icon conditions, the 'backward' presentation order slightly different icon condition, and the 'backward' presentation order dissimilar icon condition. Furthermore, it was found that the DV was not normally distributed for the 'forward' and 'backward' presentation order identical icon conditions, the 'backward' presentation order slightly different icon condition, or the 'backward' presentation order dissimilar icon condition (likely due to the outliers). Although these violations probably resulted from the outliers, the researcher did not want to remove them due to the already small sample size. Therefore, a non-parametric test was needed, and ARTool was used to analyse the data for RQ2.3. An example of ARTool's output for the accuracy data is given in Appendix-5.25.

5.2.8.4 RQ2.4: What is the impact of the degree of visual similarity between icons on the accuracy with which a participant interprets the icons?

RQ2.4 was addressed by investigating the effects of icon visual similarity on participants' accuracy at recognising an icon. The hypotheses and the null hypothesis of RQ2.4 were:

- **$H_{2.4}$:** Participants' accuracy at interpreting the meaning of an icon will be significantly different between the four icon types, in which participants will interpret icons that are identical to the known 'target' icon with the highest accuracy scores, participants will interpret slightly different icons with lower accuracy scores than the identical icons, participants will interpret mostly different icons with lower accuracy scores than the slightly different icons, and finally, participants will interpret the icons that are dissimilar to the known 'target' icon with the lowest accuracy scores.
- **H_0:** The participants' accuracy of interpretation will not be significantly different between the four icon types.

To address the hypothesis of RQ2.4, the data of the two presentation order groups ('forward' and 'backward') was combined. Furthermore, the data from the two task types (hotel and flight) was combined. The analysis for RQ2.4 focused on comparing the participants' accuracy of interpretation across the four icon types (identical, slightly different, mostly different, and dissimilar).

The initial analysis plan for RQ2.4 was to conduct repeated measures ANOVA due to the within-subjects design of the experiment. The assumptions for this ANOVA type were verified:

1. One of the assumptions of the repeated measures ANOVA is that there are no significant outliers. To test for this, the data points were graphed in boxplots and z-scores were calculated. The boxplots and z-scores indicated that for the identical icon type there was one significant outlier (z-score +/- 2.68) and one non-significant outlier (z = -1.71). The boxplot for these outliers can be seen in Appendix-5.26. Since there were only 24 participants in the study, it was decided that removing the outliers was not a good option as it would reduce the sample size.
2. Another assumption of the repeated measures ANOVA is that the distribution of the dependent variable should be approximately normally distributed. To check if the data were normally distributed, a Shapiro-Wilk normality test was run (this is the preferred test when there are less than 50 participants). The Shapiro-Wilk test showed that the data for the identical icons were not normally distributed; however, the data were normally distributed for the other three icon types (see Appendix-5.27). The histogram for the identical icons can be seen in Appendix-5.28.
3. The final assumption of the repeated measures ANOVA is that there is sphericity. To test this assumption, Mauchly's test of sphericity was conducted. Mauchly's test of sphericity was not significant, indicating that the assumption of sphericity was not violated.

An alternative approach for analysing abnormally distributed data is to use a corresponding non-parametric test. The Friedman test is the non-parametric alternative to the one-way ANOVA with repeated measures; hence, a Friedman test was used to analyse the data for RQ2.4.

5.2.8.5 RQ2.5: Is there a correlation between speed of recognition and accuracy of interpretation for each icon type (the four icon types)?

RQ2.5 was addressed by investigating the correlation between speed of recognition and accuracy of interpretation for the four icon types separately. The hypothesis and the null hypothesis of RQ2.5 were:

- **$H_{2.5}$:** For each of the four icon types, there will be a significant positive correlation between the speed of recognition and accuracy of interpretation.
- **H_0:** For each of the four icon types, there will be no significant correlation between the speed of recognition and accuracy of interpretation.

To address the hypothesis of RQ2.5, the data of the two presentation order groups (‘forward’ and ‘backward’) was combined. Furthermore, the data from the two tasks types (hotel and flight) was combined. The analysis for RQ2.5 focused on correlating the speed of participants’ recognition and their accuracy of interpretation for the four icon types (identical, slightly different, mostly different and dissimilar) independently. For RQ2.5 the within-subjects factor of icon type was analysed. In the prior analyses of Study II, it was determined that both the speed of recognition and accuracy of interpretation data were non-parametric. Taking this and the small sample size into consideration, Kendall's Tau-b was used to calculate the correlation coefficients. Kendall’s Tau performed reasonably well with sample sizes more than 10 and less than 25 (Long and Cliff, 1999).

Although no hypotheses were made about potential correlation differences between the icon types, this was also tested. A Fisher’s Z-test was conducted on the correlation coefficients of each of the following groups: 1. identical icons and slightly different icons; 2. slightly different icons and mostly different icons, and 3. mostly different icons and dissimilar.

5.3 Results

In this section, the results of Study II are presented and structured according to the relevant research questions and the research hypotheses.

5.3.1 RQ2.1: Will participants who are presented with visually similar icons in a 'forward' order recognise the icons at a different speed than participants who are presented with visually similar icons in a 'backward' order?

Following the guidance of Wobbrock *et al.* (2011), one univariate ANOVA was conducted to investigate presentation order, one univariate ANOVA was conducted to investigate icon type, and one repeated measures ANOVA was conducted for the interaction of presentation order and icon type. Submission of the AR Tool ranks to these ANOVAs revealed a significant 'main' effect of icon type, F (3, 88) = 31.74, $p < .001$. However, there was not a significant 'main' effect of presentation order, nor was there a significant interaction between presentation order and icon type.

Follow-up paired samples T-tests (suggested by Wobbrock *et al.* (2011)) were conducted to investigate the 'main' effect of icon type further. Analyses for the following three icon pairs: 1. identical and slightly different icons; 2. slightly different and mostly different icons; and 3. mostly different and dissimilar icons, were conducted. These analyses revealed a significant difference between the identical and slightly different icons ($t(23) = -5.71$, $p < .001$, 95% CI [-34.06, -15.94]), the slightly different and mostly different icons ($t(23) = -2.82$, $p = .010$, 95% CI [-19.78, -3.05]); and the mostly different and dissimilar icons ($t(23) = -4.80$, $p < .001$, 95% CI [-26.72, -10.61]). The participants' median IQR ranked scores were as follows: identical icons 16.50 (7.25 to 23.75); slightly different icons 40.50 (28.25 to 66.25); mostly different icons 54.50 (41.50 to 72.00); and dissimilar icons 80.50 (60.50 to 90.75). Figure 5-8 shows the median values of the AR Tool rank values.

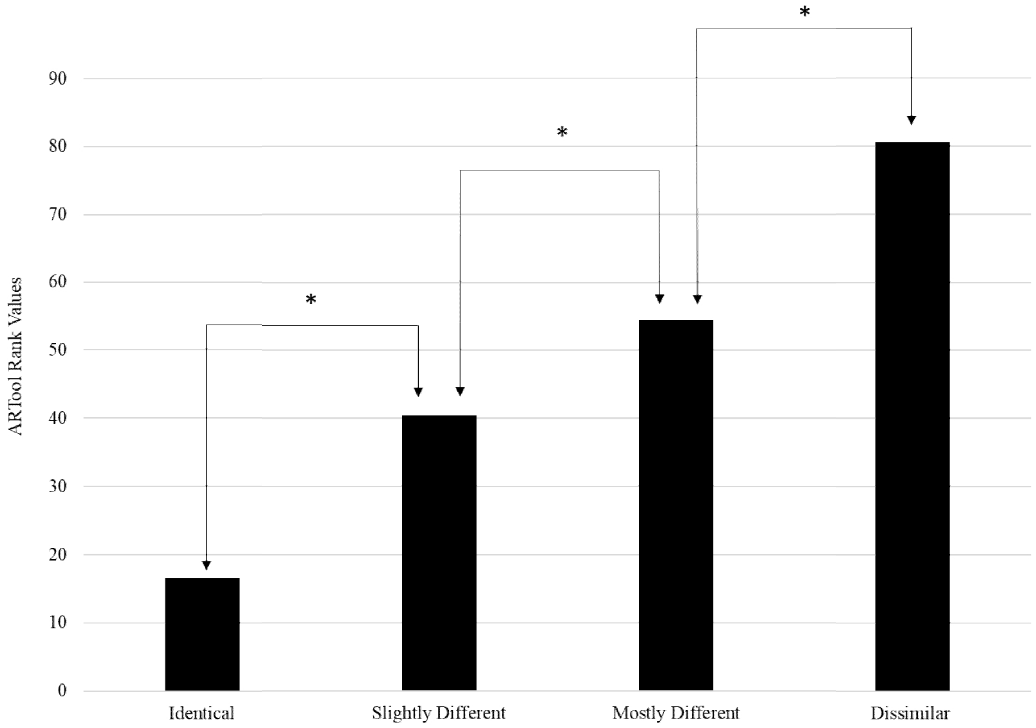


Figure 5-8: Icon type results for RQ2.1, the median AR Tool rank values for the four icon types. AR Tool rank values significantly increase as the dissimilarity between the icon types and the known 'target' icon increased. Asterisks (*) indicate statistically significant results.

5.3.2 RQ2.2: What is the impact of the degree of visual similarity between icons on the speed with which a participant recognises the icons?

There was a statistically significant difference in the participants' speed of recognition depending on icon similarity, $\chi^2(3) = 61.00$, $p < 0.001$. Post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.0125$. Median (IQR) speed of recognition for the identical, slightly different, mostly different, and dissimilar icon types were 1.07 (0.79 to 1.73), 2.989 (1.75 to 5.65), 4.06 (2.84 to 5.95), and 7.24 (4.37 to 10.52), respectively (see Figure 5-9). In general, there were significant differences in speed of recognition between the four icon types; however, speed of recognition for the slightly different and mostly different icons were not significantly different (see Table 5-3).

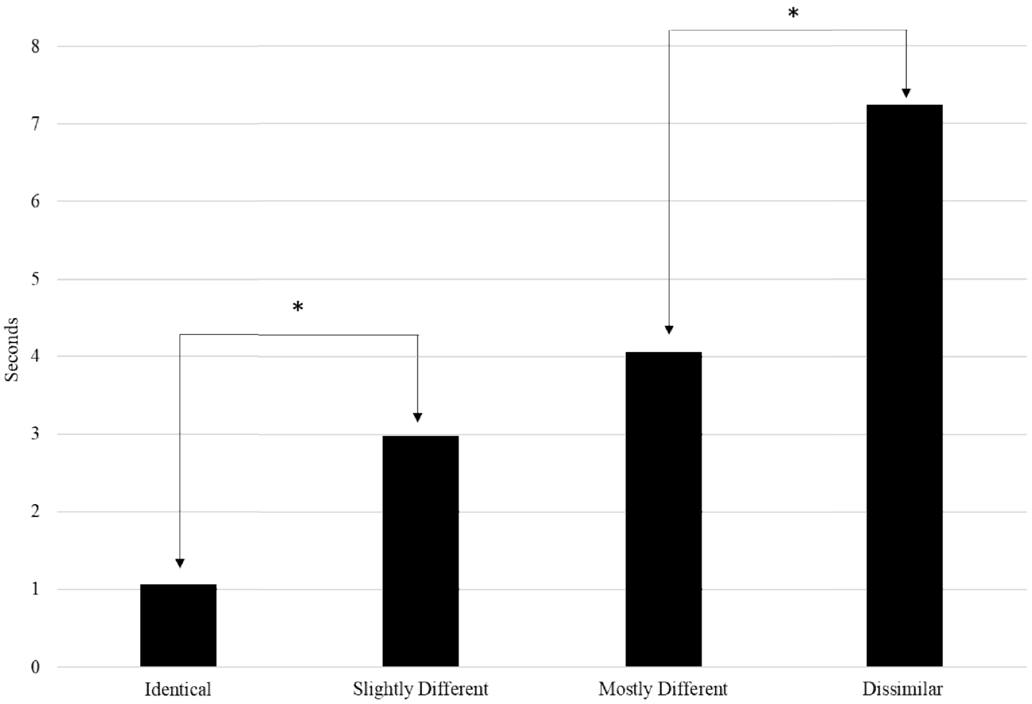


Figure 5-9: Results for RQ2.2, the median speed of recognition (measured in seconds) for the four icon types. There was a trend for participants' speed of recognition to significantly increase as the dissimilarity between the icon types and the known 'target' icon increased. Asterisks (*) indicate statistically significant results related to the hypothesis of RQ2.2.

Comparison	**Z**	**p-value (2-tailed)**
Identical and slightly different	-3.97	< .001
Slightly different and mostly different	-2.17	.030
Mostly different and dissimilar	-3.69	< .001

Table 5-3: Results of the Wilcoxon signed-rank tests with Bonferroni correction for RQ2.2. The p-value is set to $p < 0.0125$.

5.3.3 RQ2.3: Will participants who are presented with visually similar icons in a 'forward' order interpret the icons with different accuracy than participants who are presented with visually similar icons in a 'backward' order?

Following the guidance of Wobbrock *et al.* (2011), one univariate ANOVA was conducted to investigate presentation order, one univariate ANOVA was conducted to investigate icon type, and one repeated measures ANOVA was conducted for the interaction of presentation order and icon type. Submission of the ARTool ranks to these ANOVAs revealed a significant 'main' effect of icon type, F (3, 88) = 87.95, $p <$

.001. However, there was not a significant ‘main’ effect of presentation order, nor was there a significant ‘interaction’ between presentation order and icon type.

Follow-up paired samples T-tests (suggested by Wobbrock *et al.*, 2018) were conducted to investigate the ‘main’ effect of icon type further. Analyses for the following three icon pairs: 1. identical and slightly different icons; 2. slightly different and mostly different icons; and 3. mostly different and dissimilar icons, were conducted. These analyses revealed a significant difference between the identical and slightly different icons ($t(23) = 13.15$, $p < .001$, 95% CI [46.00, 63.17]), the slightly different and mostly different icons ($t(23) = -7.36$, $p < .001$, 95% CI [-33.21, -18.63]); and the mostly different and dissimilar icons ($t(23) = 7.57$, $p < .001$, 95% CI [23.07, 40.43]). The participants’ median IQR ranked scores were as follows: identical icons 83.50 (78.63 to 92.50); slightly different icons 33.50 (17.00 to 42.00); mostly different icons 58.00 (50.50 to 66.50); and dissimilar icons 21.50 (11.00 to 37.63). Figure 5-10 shows the median values of the AR Tool rank values.

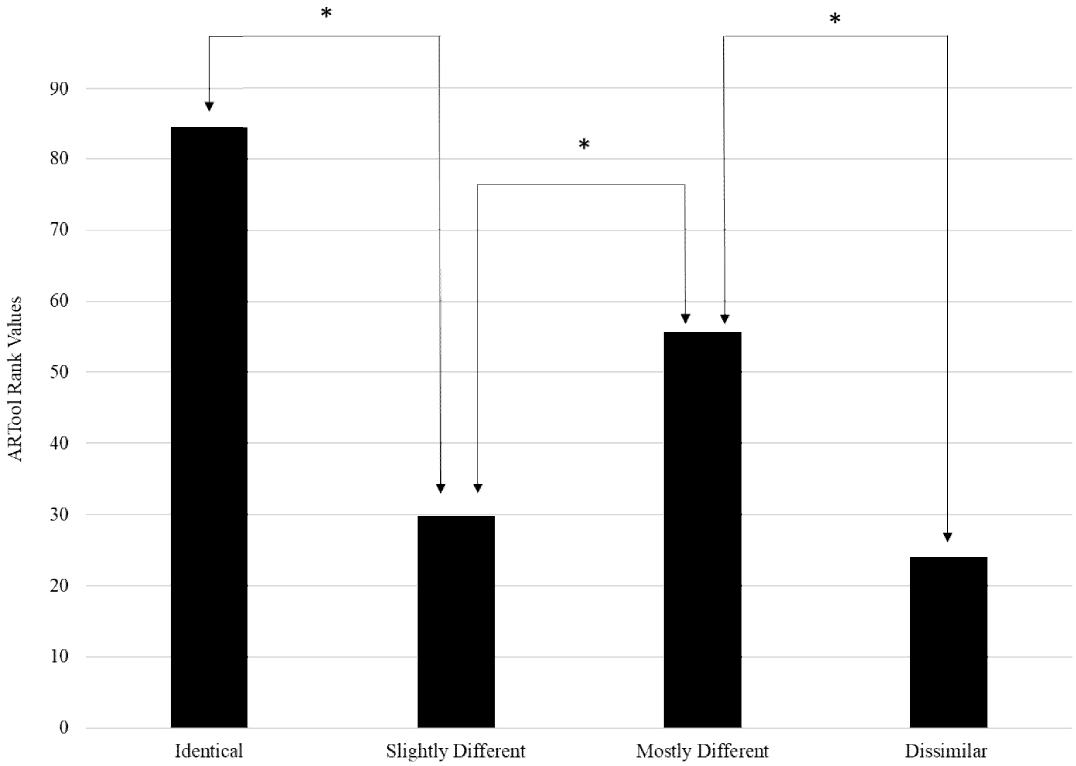


Figure 5-10: Results for RQ2.3, the median accuracy of interpretation scores (y-axis) for the four icon types. There was a trend for participants' accuracy of interpretation to significantly decrease as the dissimilarity between the icon types and the known 'target' icon increased. However, although the participants were significantly different in their abilities to accurately interpret the slightly different and mostly different icons, these icons showed a reverse pattern. Participants were more accurate at interpreting the mostly different than the slightly different icons (higher scores indicate greater accuracy). Asterisks (*) indicate statistically significant results.

5.3.4 RQ2.4: What is the impact of the degree of visual similarity between icons on the accuracy with which a participant interprets the icons?

There was a statistically significant difference in the participants' accuracy of interpretation depending on icon similarity, $\chi^2(3) = 59.00$, $p < 0.001$. Post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.0125$. Median (IQR) accuracy scores for the identical, slightly different, mostly different, and dissimilar icon types were 5.00 (4.88 to 5.00), 2.83 (2.00 to 3.13), 3.67 (3.21 to 4.29), and 2.42 (2.00 to 2.67), respectively (see Figure 5-11). There were significant differences in accuracy scores between the four icon types (see Table 5-4).

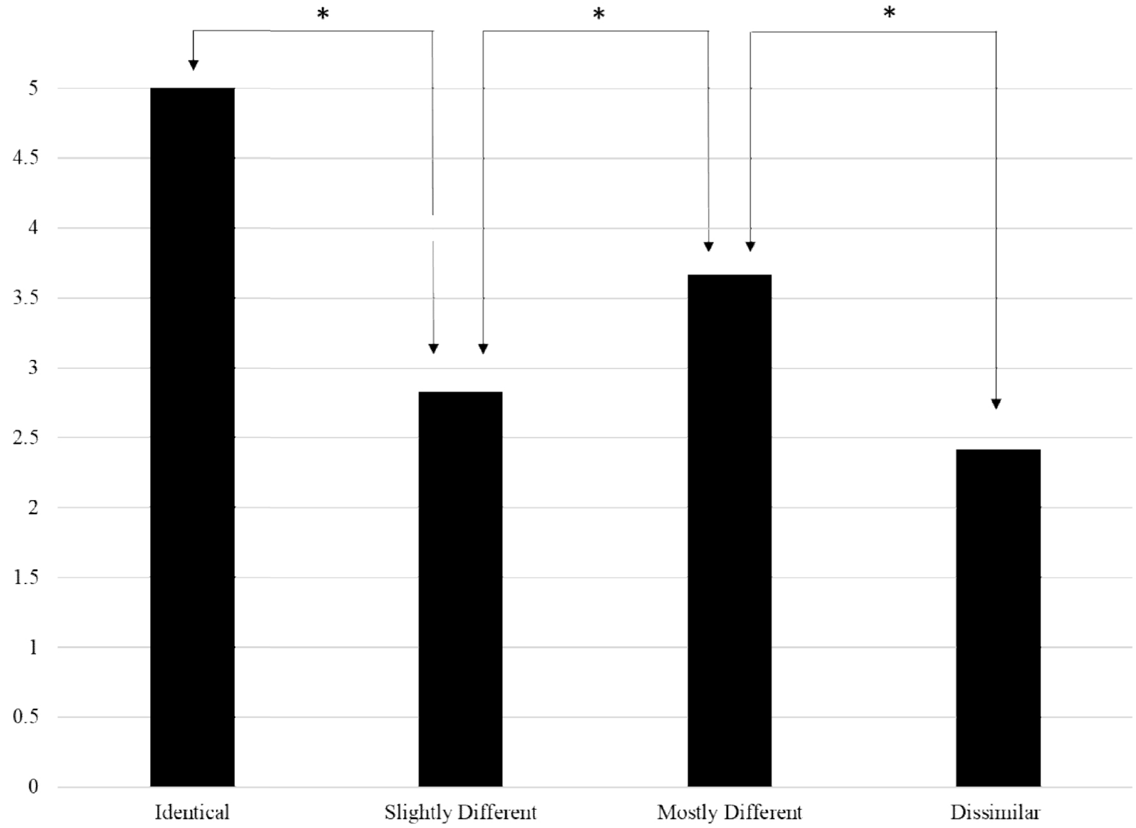


Figure 5-11: Results for RQ2.4, the median accuracy scores for the four icon types. There was a trend for participants' accuracy scores to decrease as the visual similarity between the icon types and the known 'target' icon decreased. However, while the difference between the slightly different and mostly different icons is significant, these icon types showed a reverse pattern. Participants were more accurate at interpreting the mostly different than the slightly different icons (higher scores indicate greater accuracy). Asterisks (*) indicate statistically significant results related to the hypothesis of RQ2.4.

Comparison	**Z**	**p-value (2-tailed)**
Identical and slightly different	-4.29	< .001
Slightly different and mostly different	-3.96	< .001
Mostly different and dissimilar	-4.18	< .001

Table 5-4: Results of the Wilcoxon signed-rank tests with Bonferroni correction for RQ2.4. The p-value is set to $p < 0.0125$.

5.3.5 RQ2.5: Is there a correlation between speed of recognition and accuracy of interpretation for each icon type (the four icon types)?

No significant correlations between the speed of recognition and accuracy of interpretation were found for any of the four icon types (identical, slightly different, mostly different, and dissimilar). The correlation graph for the identical icons can be seen in Figure 5-12, for the slightly different icons Figure 5-13, for the mostly different icons Figure 5-14, and for the dissimilar icons Figure 5-15. Furthermore, there were no statistically significant correlation differences for the three tested icon groups: 1.

identical icons and slightly different icons; 2. slightly different icons and mostly different icons, and 3. mostly different icons and dissimilar icons.

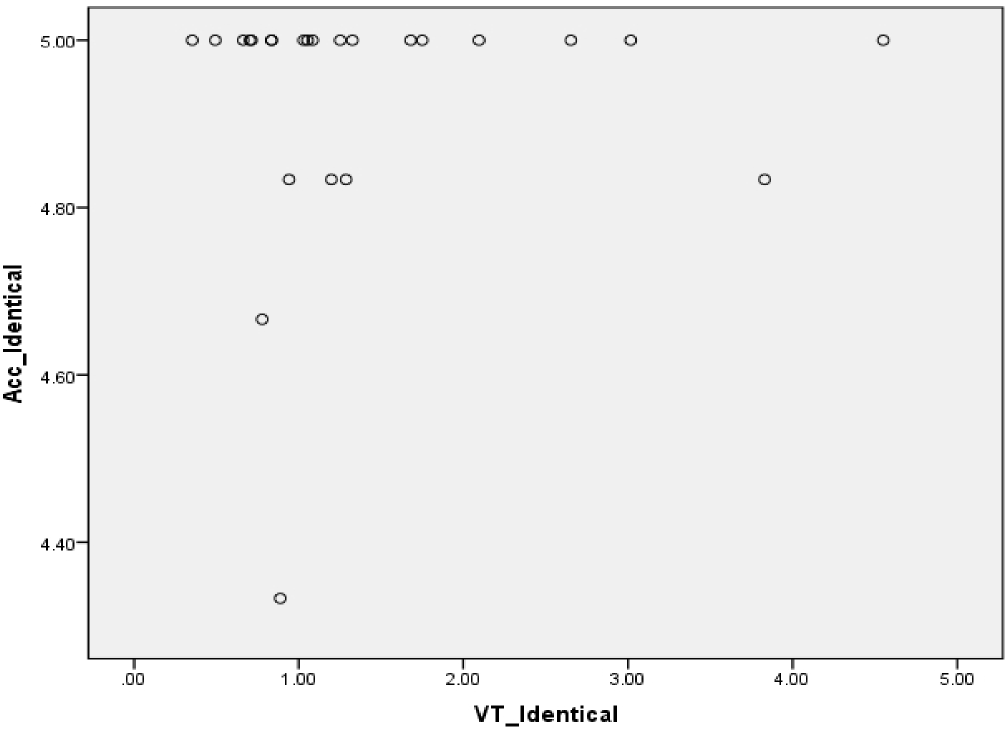


Figure 5-12: Graph for RQ2.5 showing the correlation between viewing time/speed of recognition (x-axis) and accuracy of interpretation (y-axis) for the identical icons.

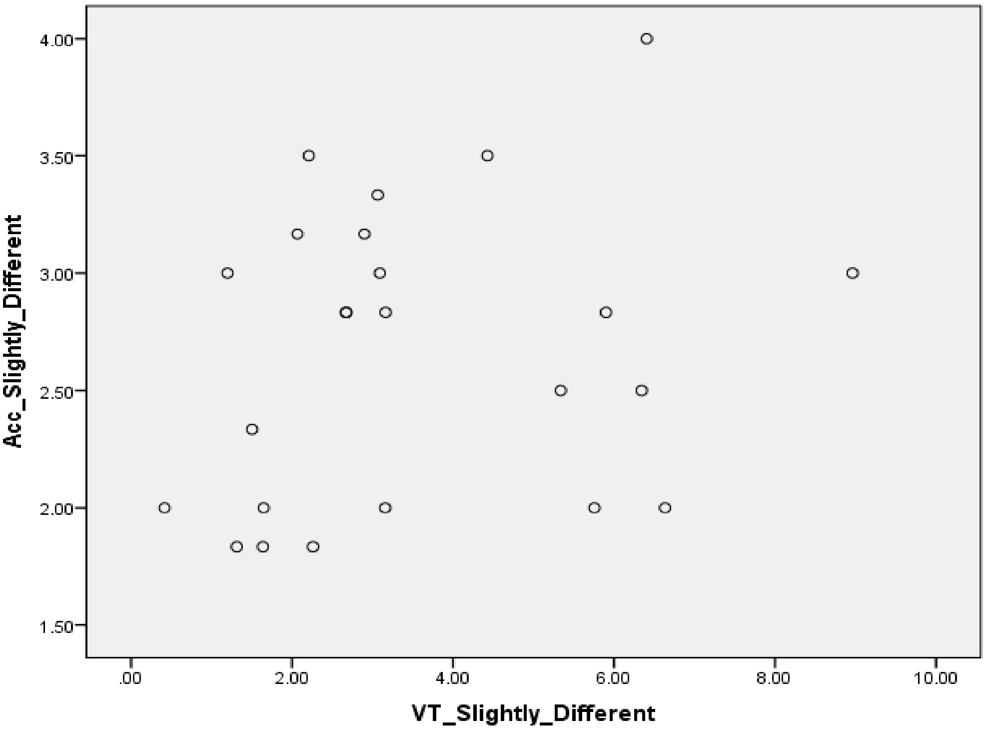


Figure 5-13: Graph for RQ2.5 showing the correlation between viewing time/speed of recognition (x-axis) and accuracy of interpretation (y-axis) for the slightly different icons.

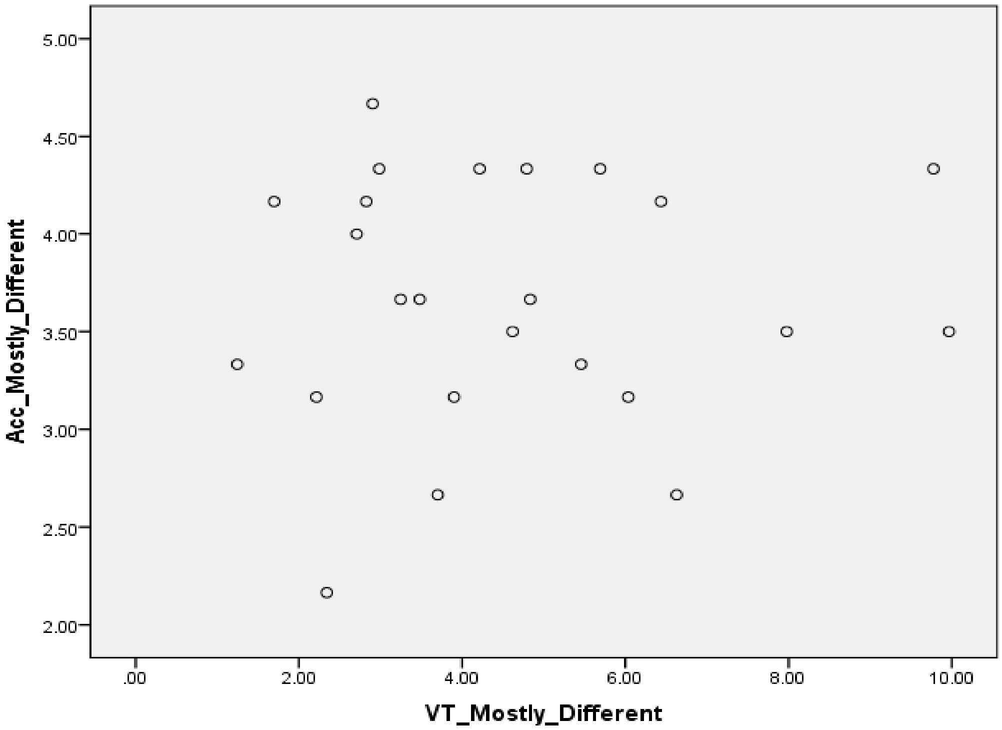


Figure 5-14: Graph for RQ2.5 showing the correlation between viewing time/speed of recognition (x-axis) and accuracy of interpretation (y-axis) for the mostly different icons.

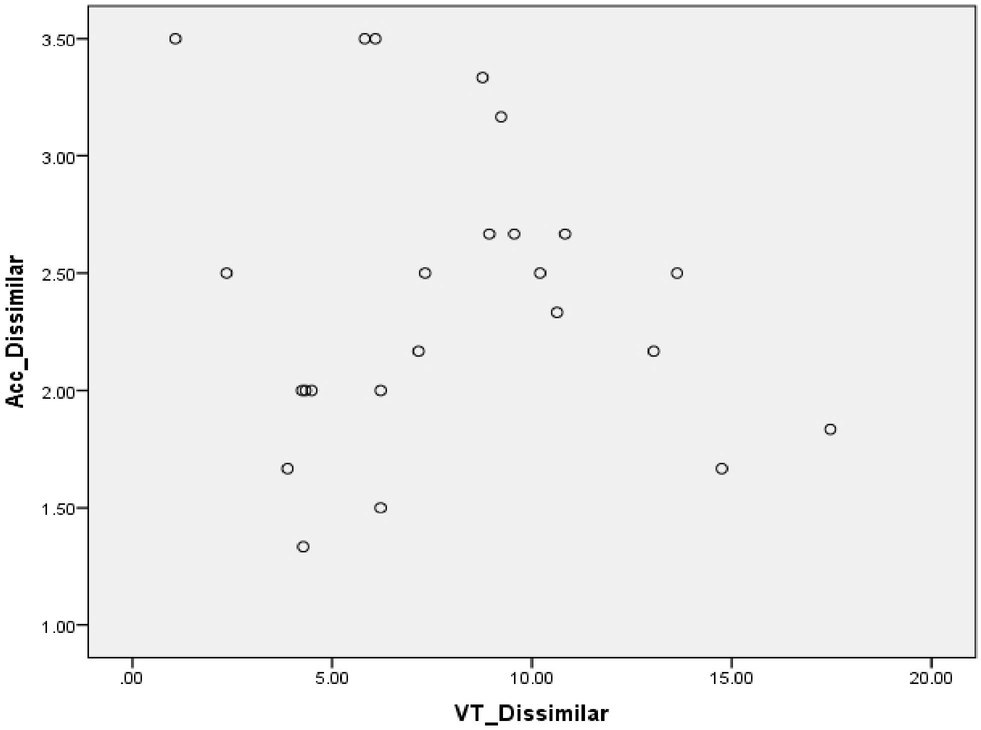


Figure 5-15: Graph for RQ2.5 showing the correlation between viewing time/speed of recognition (x-axis) and accuracy of interpretation (y-axis) for the dissimilar icons.

5.4 Discussion

The hypothesis of RQ2.1, that presentation order would influence the participants' speed of recognition, was not supported. However, it seems that icon type does affect the participants' speed of recognition. The analysis for RQ2.1 was based on the rank scores created by the AR Tool program; therefore, the results of these analyses were based on the rank scores and not the raw speed of recognition data. To further investigate the possible effects of icon type on participants' speed of recognition, RQ2.2 was established. The analysis for RQ2.2 was conducted on the raw (not AR Tool ranked) speed of recognition data.

The hypothesis of RQ2.2 was largely supported. The participants' speed of recognition did decrease as the visual similarity between the icon types and the known 'target' icon decreased. However, the decrease in the speed of recognition between the mostly different and slightly different icons was not statistically significant. These results may indicate that these two icons types were too similar, and therefore, participants did not need a significantly different amount of time to recognise the icons.

Prior research has shown that fixation equates to attention (Jabeen, 2010) and cognitive processing (Griffin *et al.*, 2004). Typically, slower viewing times mean that a participant is paying more attention to stimuli, which is thought to equate to enhanced cognitive processing. In this experiment, slower viewing times reflected the amount of time it took participants to decide the meaning of the icon. As such, viewing time can be used as a measure of participants' speed of recognition. The results of this experiment indicate that there was a trend for participants' speed of recognition to decrease as icon similarity decreased, which may indicate that as the similarity between icons decreases, participants require more attention and cognitive processes to identify icons.

Although this trend was present, the participants' speed of recognition was not significantly different for the slightly different and mostly different icons. Prior research has indicated that simple icons elicit quicker responses than complex icons (McDougall *et al.* 2000; Schröder and Ziefle, 2008). Furthermore, prior research has also shown that icons that are more concrete and have a smaller articulated distance require shorter reaction times and less cognitive processing than less concrete icons with a larger articulated distance (Rogers, 1986; Blankenberger and Hahn, 1991; Cheng and Patterson, 2007). Therefore, the lack of a significant difference between the slightly different and mostly different icons may indicate that the icons were too similar. Icon design is a complex activity and it is plausible that either prior participant experience or subtle misjudgements were at play here.

Furthermore, the analysis for RQ2.2 consisted of combining the data from the two task types (hotel and flight booking) and combining the data from the presentation orders (‘forward’ and ‘backward’). The task types had six completely different icon families. Furthermore, the ‘forward’ and ‘backward’ groups viewed different slightly different and dissimilar icons; as well as these icons being presented in different orders to the two groups. Hence, these unexpected patterns in the speed of recognition may be related to presentation order and design of the icons.

The hypothesis of RQ2.3, that presentation order would influence the participants’ abilities to accurately interpret the icons, was not supported. However, it seems that icon type does affect the participants’ abilities to accurately interpret the icons. The analysis for RQ2.3 was based on the rank scores created by the AR Tool program; therefore, the results of these analyses were based on the rank scores and not the raw accuracy data. To further investigate the possible effects of icon type on participants’ accuracy, RQ2.4 was established. The analysis for RQ2.4 was conducted on the raw (not AR Tool ranked) accuracy data.

The hypothesis of RQ2.4 was largely supported. The participants’ accuracy of interpretation showed a trend to decrease as the visual similarity between the icon types and the known ‘target’ icon decreased. However, this trend was not true for the slightly different and mostly different icons. Participants were significantly less accurate at interpreting the meaning of the slightly different icons than they were the mostly different icons. The analysis for RQ2.4 consisted of combining the data from the task types (hotel and flight booking) and combining the data from the presentation orders (‘forward’ and ‘backward’). The task types had six completely different icon families. Furthermore, the ‘forward’ and ‘backward’ groups viewed different slightly different and dissimilar icons; as well as these icons being presented in different orders to the two groups. Hence, these unexpected patterns in accuracy may be related to presentation order and design of the icons.

The hypothesis for RQ2.5 stated, for each of the four icon types, there will be a significant positive correlation between speed of recognition and accuracy of interpretation. The hypothesis for RQ2.5 was not supported. There were no significant correlations between the speed of recognition and accuracy of interpretation for any of the four icon types. It may be the case that no significant results were found for these correlations because there was not a strong monotonic relationship between the variables. Kendall's tau-b determines whether there is a monotonic relationship between two variables; however, the presence of a monotonic relationship is not an assumption of the Kendall's tau-b analysis method. Finally, no statistically significant correlation differences for the three tested icon groups: 1. identical icons and slightly different icons; 2. slightly different icons and mostly different icons, and 3. mostly different icons and dissimilar icons, were found. Typically, in psychology, as peoples' response times increase so does their accuracy, known as the speed-accuracy trade-off (Zimmerman, 2011). In the current experiment, the speed of recognition was analogous to response times. Hence, in the current experiment, when a participant took longer to recognise an icon, they may have also been more accurate at interpreting the icon. It was expected that this would be true for each of the four icon types. Hence, no hypothesis was made about potential correlation differences; and the lack of significance for these tests may support the presence of the speed-accuracy trade-off for all four icon types.

So far, the discussion has focused on the hypotheses of Study II; however, additional data were collected during Study II (although no hypotheses were made). During Study II, data about the participants' reasons for believing an icon would behave a certain way was also collected. Data about participants' reasoning was collected when they answered the question, 'Why do you think that will happen?' Inductive coding was used to create nine categories of reasoning: 1. shape, 2. specific previous website (external to the study), 3. unspecific previous website (external to the study), 4. internal to the study, 5. text inside the icon, 6. internal context within the icon, 7. task-related, 8. real-world analogy, and 9. icon location. These reasons are defined as:

1. **Shape:** refers to responses where the participant attributed their interpretation to the shape of the icon.
2. **Specific previous website (external to the study):** refers to responses where the participant attributed their interpretation to an element that was experienced before the experiment with naming a specific website. (i.e. 'I've seen it before on other websites, like Facebook.').
3. **Unspecific previous website (external to the study):** refers to responses where the participant attributed their interpretation to an element that was experienced before the experiment without naming a specific website. (i.e. 'it shows three stars and usually stars mean rating. I've seen it before in travel websites').
4. **Internal to the study:** refers to responses where the participant attributed their interpretation to an element that was experienced within the experiment. (i.e. 'from the previous coaching').

5. **Text inside the icon:** refers to responses where the participant attributed their interpretation to the text represented in the icon. (i.e. ‘because of the VIP lettering on the icon’).
6. **Internal context within the icon:** refers to responses where the participant attributed their interpretation to elements within the icon. (i.e. ‘because the P is consistent with the previous two icons as well as the island icon and the P being within the island.).
7. **Task-related:** refers to responses where the participant attributed their interpretation to the undertaken process/task (i.e. ‘because of the positioning of that icon as well as the processes we are going through’).
8. **Real-world analogy:** refers to responses where the participant attributed similarity or likeness between the visual representation of an icon and a real-world object. (i.e. ‘Because of the heart and people use a heart as an emoji to say how much they like things.’).
9. **Icon location:** refers to responses where the participant attributed their interpretation to the location of the icon in the interface. (i.e. ‘because it’s next to the money on your account, so if you wanted to put money into your account.’). From these categories, shape was the most common reason participants thought an icon would perform in a certain way - shape was cited 284 times. Interval to the study was the second most cited category, cited 250 times. For how often each of the nine categories was cited, refer to Table 5-5. Hence, it seems that the visual appearance (shape) of the icons and their internal use in the study were the biggest drivers of how participants interpreted the function of the icons.

Reason	*Frequency*
Shape	284
Internal to the study	250
Unspecific previous website (external to the study)	71
Internal context within the icon	70
Real-world analogy	63
Text inside the icon	55
Specific previous website (external to the study)	7
Icon location	3
Task-related	1

Table 5-5: Responses to 'Why do you think that will happen?' and how often these responses were given.

Work by Kahneman (2011) suggests that two systems can be used when people make judgments and choices. The first system (fast thinking) is fast and automatic, and the second system (slow thinking) is slower and more deliberate. The first system, or fast thinking, requires almost no effort from the person and the person has no feeling of control over this process. This type of thinking often occurs when judgments and choices that are made frequently must be made. For example, when asked what 2+2 equals, a person will answer 4 without giving the answer much or any thought - this is fast thinking. The second system, or slow thinking, requires effort and concentration. A person must be consciously aware when utilising slow thinking. This system is used for monitoring one's behaviour in certain situations, for example monitoring your walking pace.

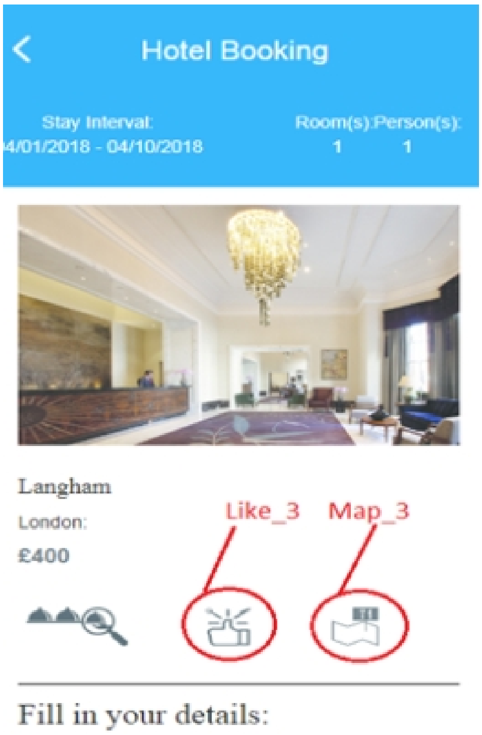
While these two systems have distinct roles, they can also work together. System one answers questions and runs less cognitively taxing processes to conserve energy for the second system to have cognitive resources to solve more complex problems. However, sometimes system one does not have all the answers and may rely on system two for more information. Kahneman (2011) suggests that system two is designed to monitor the thoughts and actions that system one promotes. System two can also control the thoughts and actions of system one by encouraging, suppressing, or modifying behaviours.

System one operates best when the world is predictable, straight-forward, and easy to understand, hence system one likes to interpret the world in a way that is stable and predictable. However, this vision of predictability is an illusion. The brain tries to predict the future based on the past; however, according to Kahneman (2011) humans do not understand the past as much as they think they do.

System one uses memory to make future predictions, which can drive choices and decisions. However, the memories of humans are known to be fallible, for example, the brain more strongly remembers events which have a strong emotional attachment. For example, if one has a painful divorce, they are likely to remember the marriage as painful, although there were surely good moments. These long-term emotional memories are one-way choices and decisions can be influenced. However, Kahneman (2011) also suggests that short-term memory, which is only 15 minutes long, can also be influenced by the past and emotions, which may impact decision-making.

This work by Kahneman (2011) is relevant for the findings of Study II because it suggests that people make minor decisions automatically without cognitive thought and that these decisions may be influenced by both short-term and long-term memory. Hence, participants in Study II may have made judgments about the icons automatically without cognitive thought. These types of judgments may have been influenced by the participants' previous experiences in long-term memory (e.g., Facebook icons similar to the icons in the current study) or even by short-term memory of the icons in the session they attended. Since participants cited 'internal to the study' as the second leading reason why they thought that an icon would behave in a certain way, it seems that their short-term memory of the icons which were used early in the experiment influenced how they interpreted the icons that were later used in the experiment.

Another interesting observation in this study was that for the *Like_3* and *Map_3* icons, which were mostly different from the known 'target' icon, the participants moderately to accurately interpreted the icons (see Figure 5-16). For *Map_3*, the participants were mainly accurate and for *Like_3*, they moderately to accurately interpreted the icon. Participants may have been able to accurately interpret these icons even though they were mostly different to the known 'target' icon because the Map_3 and Like_3 icons share common shapes (fork and knife (cutlery) often signify a restaurant). This is consistent with the findings in Shen *et al.*, (2018), who stressed that users are familiar with the icon if they were familiar with the objects depicted in the icon, which would affect the users' performance in identifying the icons.



Hotel Booking

Stay Interval:
4/01/2018 - 04/10/2018

Room(s):Person(s):
1 1

Langham
London:
£400

Fill in your details:

Figure A: Accuracy of Participants' Interpretations of Map_3

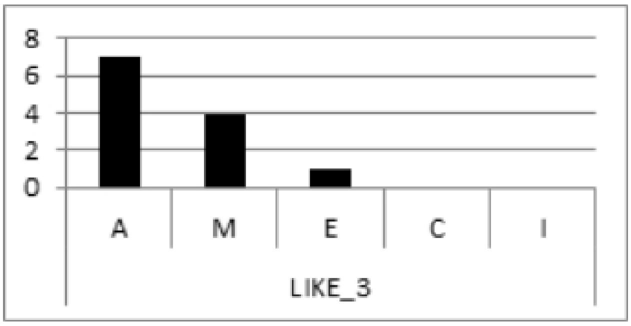


Figure B: Accuracy of Participants' Interpretations of Like_3

Figure 5-16: The Like_3 and Map_3 icons are shown on the left-hand side; they are circled in red and labelled. Participants' accuracy of interpretation for the Map_3 icon is shown in Figure A (upper right-hand corner). Participants' accuracy of interpretation for the Like_3 icon is shown in Figure B (lower right-hand corner).

Considering these findings, the following suggestions are given for future studies. For future studies, it may be beneficial to have a larger sample size. The viewing times and accuracies tend not to be normally distributed for this small group. Although having a larger sample size might not lead to a normal distribution of accuracy for the identical icons, as was noted earlier, participants tended to be accurate at interpreting the identical icons. Furthermore, in the future, it may be beneficial to use a larger sample size so that the four groups have more participants and can be compared. Additionally, for the research questions in this study, the data from the two task types (hotel booking and flight booking) were combined. However, these two tasks utilised different icons and any possible effects that arose from these differences were ignored. For RQ2.1, RQ2.4, and RQ2.5, the data from the two presentation order groups ('forward' and 'backward') was also combined. However, the presentation order groups were presented with different icons for the slightly different and dissimilar icon types. In the future, it may be beneficial to create an experiment which uses icons that are visually the same for different presentation orders.

5.5 Conclusion

From the results presented here, there was a trend for participants' speed of recognition to significantly increase as the dissimilarity between the icon types and the known 'target' icon increased. However, the difference between the slightly and mostly different icon types was not statistically significant. Furthermore, there was a trend for participants' accuracy of interpretation to significantly decrease as the dissimilarity between the icon types and the known 'target' icon increased. However, participants were significantly more accurate at interpreting the meaning of the mostly different icons than the slightly different icons (which was unexpected). These results may indicate that participants need more attention and cognitive processes to determine the semantic meaning of icons that differ from a known 'target' icon. Finally, no significant correlations were found between the speed of recognition and accuracy of interpretation for any of the four icon types.

To conclude, the results of this study indicate that: (1) As icon dissimilarity increases so does the speed of recognition, and (2) identical icons lead to higher accuracy, while in general dissimilar icons lead to lower accuracy. It is suggested that when designing icons, designers should use icons that are familiar to the user so that the user will quickly and accurately interpret the meaning of the icon.

It was also concluded from Study II that participants were unexpectedly accurate on some of the slightly different, mostly different, and dissimilar icons (for example the Like_3 and Map_3 icons). These results, along with the fact that participants cited internal to the study as their reason for believing an icon would act in a certain way 250 times, suggests that participants may have become familiar with the icons. This had an influence on icon recognition and understanding. Hence, Study III will ask the question, what is the impact of the visual context (image presented with icon versus icon only) on participants' interpretations of the icons? Study III will not use repetitive icons. The participants should not become familiar with the icons and therefore, their interpretations should not be influenced by their familiarity with the icons. Furthermore, Study III will have a *'visual context'* group and a *'no visual context'* group. The 'visual context' group will be shown icons within their intended context. The 'no visual context' group will be shown the same icons; however, no contextual information was given.

Chapter 6 Exploring the Impact of Visual Context on Participants' Interpretations of Icons (Study III)

6.1 Motivation and Research Questions

In Study II, some of the participants were more accurate at identifying icons than expected. For instance, there was a high number of accurate responses for the 'mostly different' icons. A mostly different icon was one where two or more fractions of the icon were different from the known 'target' icon, although the mostly different icon also shared one or more fractions with the known 'target' icon. In Study II, it was expected that participants would not be very accurate at interpreting the meaning of the mostly different icons, however the participants were able to accurately identify the mostly different icons. This result was unexpected and therefore was investigated further.

Furthermore, in Study II it was observed that the participants used the visual context in which the icons were presented to interpret their meaning (this was especially true for the slightly different, mostly different, and completely different icons). This observation came from both the verbal responses of the participants and the eye-tracking data. For example, Table 6-1 shows the response of a participant when they were asked why (the reason) they interpreted the meaning of an icon in the way they did. This participant relied on the surroundings of the icon to accurately interpret its function. Figure 6-1 shows the gaze of a participant; the participant not only looked at the icon, but also its surrounding visual context. Based on such observations, it was concluded that participants might have referred to the visual context of the icon to interpret the icon's meaning.

Participant id	Icon type	Icon	Reason
U2	Completely Diff		Because it's next to the money on your account, so if you wanted to put money into your account.

Table 6-1: A participant who relied on the context of the icon to interpret its meaning.

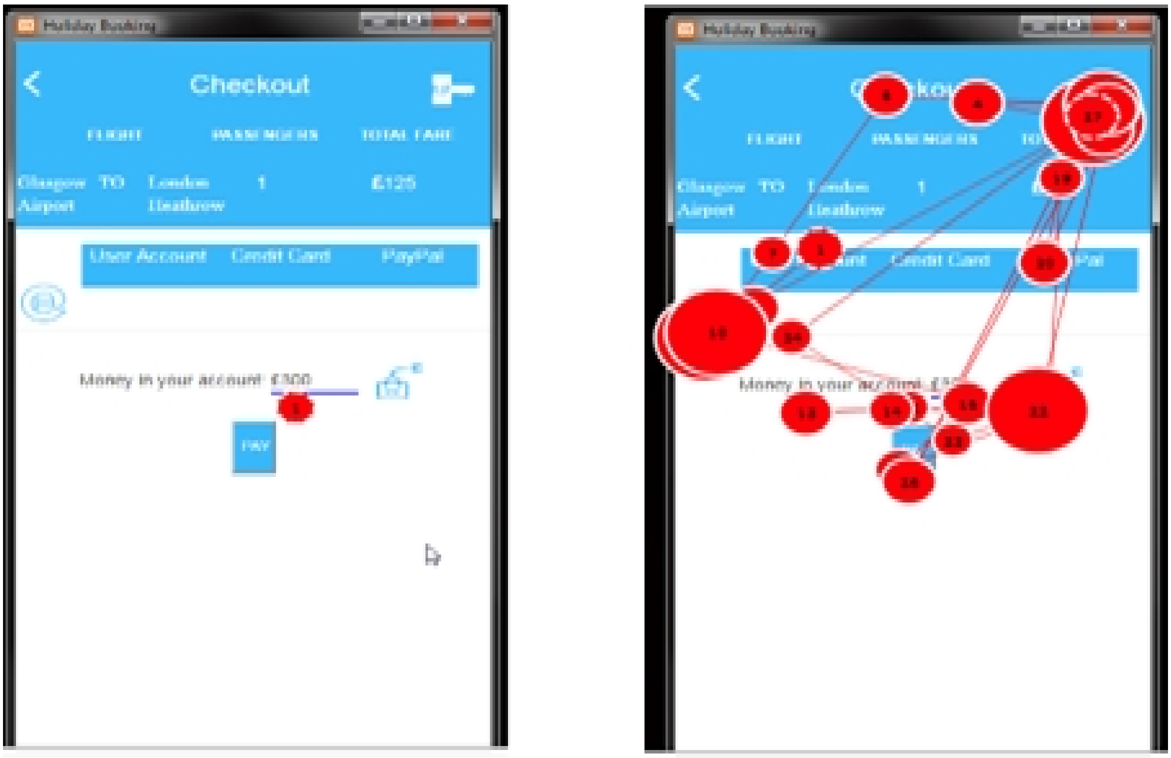


Figure 6-1: On the left is a screenshot of a web page before a participant started the free viewing phase. On the right is a screenshot of a participant's gaze during the viewing phase and before interpreting the icon's meaning. Icons were dissimilar and they were presented during a flight booking task in a 'forward' presentation order.

The observation that participants relied on the visual context of an icon to interpret its meaning is not surprising, as prior research has shown that when participants are presented with unfamiliar icons, in icon-only interfaces, they are no better at understanding the meaning of the icon than if a text-only interface was used (Wiedenbeck, 1999). Furthermore, Huang and Bias (2012) have shown that participants who are not familiar with icons and are not given any context to understand the icons are inaccurate and inefficient at recognising icons. Huang and Bias (2012) also showed that participants required more time and made more mistakes when interpreting icons in comparison to textual information. Huang, Bias, and Schnyer (2015) have also shown with a functional magnetic resonance imaging (fMRI) study that while icons do stimulate the semantic system, they are not cognitively processed as logographical words. Instead, icons are processed more as images or pictures. As such, the authors concluded that icons are not as efficient as words in conveying meanings.

In studies I and II, it was observed that the participants used the visual context where the icons were presented in to interpret their meaning. Therefore, this led to the formation of RQ3. Hence, Study III builds on studies I and II by comparing how a group that is provided with icons in context and a group that is not provided with icons in context differ in their ability to interpret the meaning of icons. It was hypothesised that by providing the visual context for the icons, participants will be able to understand and interpret the meaning of the icons more accurately than participants who are given no visual context. Note, in Study III the context of an icon was defined as a visual image of a web page that an icon appeared within to do an intended function. As such, Study III addressed the following research question:

- *RQ3: What is the impact of visual context on participants' interpretations of icons?*

A pilot study was conducted with two participants to test how the online study would run and how participants would interpret the icons. In the pilot study, the link to the Qualtrics form was directly sent to participants via email.

6.2 Methods

6.2.1 Study Design

An internet-based study was conducted to address RQ3, the research question of Study III. A between-subjects design was used with two groups. Half of the participants (15) were placed in the 'visual context group'. Participants in this group were shown icons in visual context before they were asked to interpret/type in the meaning of the icons. The other half of the participants were placed in the 'no visual context group'. Participants in this group were NOT shown visual context with the icons. They were shown the icons alone before they were asked to interpret/type in the meaning of the icons. In this study, the IV was the visual context and had two levels: visual context and no-visual context. The DV was the accuracy of interpretation. The hypothesis and the null hypothesis of RQ3 were:

- **H_3:** Participants who are provided with visual context when interpreting icons will be significantly more accurate at interpreting the icons than participants who are not provided with visual context.
- **H_0:** (null hypothesis) Participants who are provided with visual context when interpreting icons will not be significantly different at accurately interpreting the icons than participants who are not provided with visual context.

Qualitative data were collected in the form of typed-in textual responses after the participant viewed the icons. The data were then coded in order to create quantitative data which was analysed. These data were collected to determine the participants' interpretations of icons. The participants' accuracy of interpretation was then scored based on a five-point accuracy scale. The two groups were compared on their levels of accuracy of interpretation for the icons. The research process for Study III is illustrated in Figure 6-2.

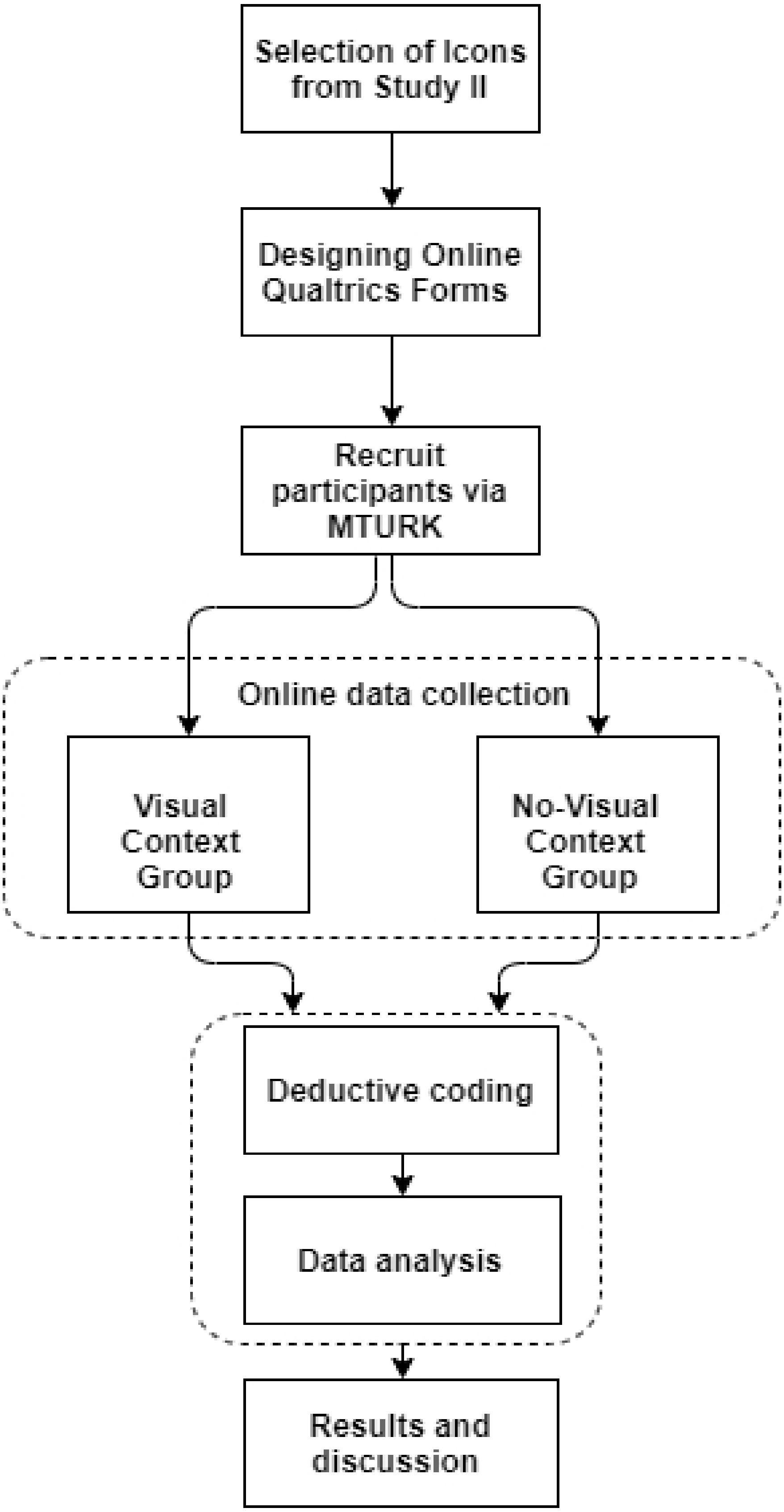


Figure 6-2: Research process for Study III.

6.2.2 Participants

A convenience sample of 30 participants was recruited via Amazon Mechanical Turk (Mturk) and asked to complete the study via Qualtrics. The 30 participants were divided evenly into two groups, the ‘visual context’ group and the ‘no visual context’ group. To participate, the participants needed to be aware of technology and have experience interacting with icons. Furthermore, the participants were required to be at least 18 years old. Mturk was set up so that a participant could only participate in the experiment once. This was to make sure that each group had independent observations. Participants were paid $2 for completing the experiment.

6.2.3 Materials

6.2.3.1 Icon Selection Process

In Study II, it was expected that the accuracy of participants’ responses would decrease as the similarity between the known ‘target’ icon and the similar/dissimilar icons decreased. However, for some icons, the participants were unexpectedly accurate at interpreting their meaning. The icons that the participants were unexpectedly accurate at interpreting fell within the slightly different, mostly different, and completely different icons types (the specific icons are listed in Table 6-2). The slightly different, mostly different, and completely different icons types, which the participants were unexpectedly accurate at interpreting, were further analysed for Study III. However, icons in the identical similarity level/type were not further analysed, as it was expected that participants would be accurate when interpreting these icons. To determine the top eight icons which participants were unexpectedly accurate at interpreting, accuracy means were calculated for icons of slightly different, mostly different, and dissimilar types in Study II and they were rounded to the nearest integer (see Appendices 6.1, 6.2 and 6.3). Appendix-6.1 shows the accuracy means for the slightly different icons in Study II. Appendix-6.2 shows the accuracy means for the mostly different icons in Study II. Appendix-6.3 shows the accuracy means for the completely different icons in Study II. The calculation of the accuracy means resulted in the selection of eight icons to be used in Study III (see Table 6-2).

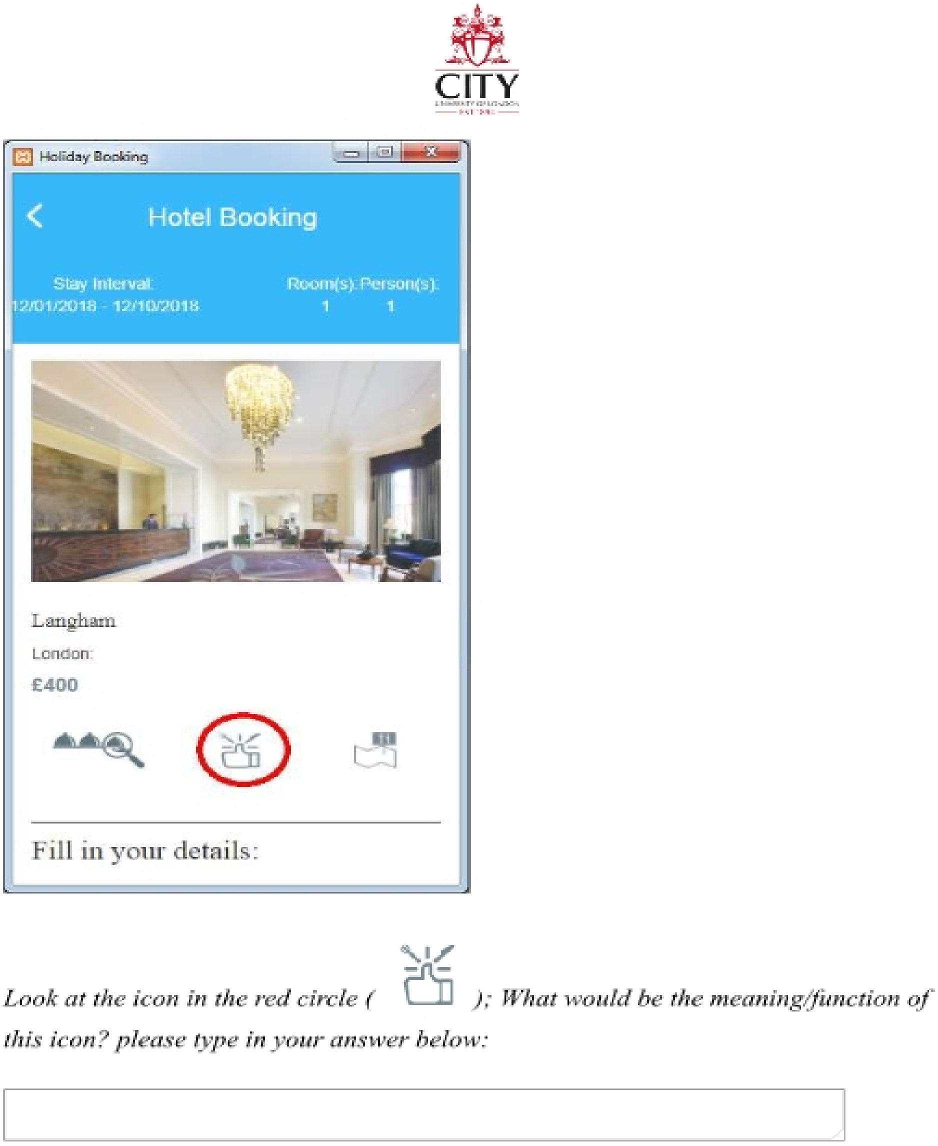
ID	*Name*	Icon	Family	Icon Type, Order	Acc Mean	Task	Meaning
1	*Email_2_FR*		Email	Slightly diff, Forward	1.91	Flight	*To ask for the discount code via e-mail*
2	*Password_2_FR*		Password	Slightly diff, Forward	1.83	Flight	*To show the secret flight offer for Today*
3	*Map_2_BK*		Map	Slightly diff, Backward	2.166	Hotel	*To show the route between the hotel and the city centre on the map*
4	*Like_3*		Like	Mostly diff	4.38	Hotel	*To show the number of customers who liked the food provided in the hotel*
5	*Map_3*		Map	Mostly diff	4.88	Hotel	*To show the nearest restaurants to the hotel on the map*
6	*UserAcc_4_FR*		User Account	Completely diff, Forward	3.83	Flight	*To add money to the user's account*
7	*Review_4_FR*		Review	Completely diff, Forward	3.16	Hotel	*To show customers' reviews of the staff in the hotel*
8	*Like_4_FR*		Like	Completely diff, Forward	2.58	Hotel	*To show the number of customers who liked the cleanliness of the hotel*

Table 6-2: The eight icons used in Study III.

6.2.4 Procedure

In Study III, the data were collected online via Mturk. Before beginning the experiment, the participants signed an online consent form (Appendix-6.5). They were then given instructions for the experiment and shown an example icon. Both groups saw the same example icon on the instruction page, but the 'visual context' group saw the example icon in a picture, whereas the 'no visual context' group saw the example icon on its own. Participants were asked, 'What would be the meaning/function of this icon?' and they were required to type in their answer to this question (see Appendix-6.4). Participants then viewed the eight icons in sequential order, however, the order of the icons was randomised (using the Qualtrics randomiser) across the participants. The 'visual context' group was shown the icons in their visual contexts (see Figure 6-3). The ‘no visual context’ group was not given any visual contextual information for the icons (see Figure 6-4). All participants could view the icons with no time limit while they typed their interpretation of the meaning of the icons into a text box. Participants were not required to click on the icons, they only needed to provide their interpretation of the icons.

After conducting the pilot study, in the real study, the link to the Qualtrics forms was placed on the Mturk page. When a participant completed the study, they received a code which they could use to be paid.



Look at the icon in the red circle ( ); What would be the meaning/function of this icon? please type in your answer below:

Figure 6-3: An example of how the participants in the ‘visual context’ group were presented with the icons and prompted for their interpretation of the meaning of the icons as they navigated through the Qualtrics’ pages.

What would be the meaning/function of this icon? Please type in your answer below:

Figure 6-4: An example of how the participants in the ‘no visual context’ group were presented with the icons and prompted for their interpretation of the meaning of the icons as they navigated through the Qualtrics’ pages. In this example, the participant would see the screen with the Like icon, and type in their interpretation of the meaning of the icon. Then they would click to go to the next screen.

6.2.5 Data Analysis

Participants’ responses were rated by an expert who did not collect the data, on a five-point scale: accurate = 5, moderate = 4, conflicting = 3, erroneous = 2, and incapable = 1 (Islam and Bouwan, 2016). Accurate meant that the participant’s interpretation matched the designed meaning. Moderate meant that the participant’s interpretation matched the design meaning to some extent, but their interpretation did not completely match the design meaning. Conflicting meant that the participant tried to get the meaning by saying different interpretations, but in the end, they were not correct. Erroneous meant that the participant’s interpretation was completely wrong. Incapable meant that the participant did not attempt to provide a meaning; in these cases, the participants answered ‘I don’t know’ or a variant of this statement. Table 6-3 shows the accuracy of participants’ interpretations of the tested icons in ‘visual context’ and ‘no visual context’ groups. It also shows the accuracy mean for each participant across all the icons, and Table 6-4 shows the descriptive statistics of the accuracy scores for the two groups.

ID	1	2	3	4	5	6	7	8	Mean
	Email_2_FR	Password_2_FR	Map_2_BK	Like_3	Map_3	UserAcc_4_FR	Review_4_FR	Like_4_FR	
1C	1	2	3	3	5	5	5	4	3.5
2C	4	2	3	3	5	5	2	2	3.25
3C	1	1	5	4	5	5	1	3	3.125
4C	3	2	2	2	5	3	2	2	2.625
5C	4	2	2	2	3	5	2	2	2.75
6C	5	1	4	2	3	5	2	2	3
7C	4	2	3	2	3	2	2	2	2.5
8C	2	2	3	3	3	4	2	2	2.625
9C	3	2	4	2	5	5	2	2	3.125
10C	3	2	2	4	5	5	2	1	3
11C	3	2	3	3	3	5	2	2	2.875
12C	1	2	4	3	3	3	2	3	2.625
13C	4	2	3	3	3	2	2	2	2.625
14C	3	2	3	4	3	5	2	2	3
15C	5	2	3	5	5	5	2	3	3.75
1NC	4	2	1	4	5	5	2	1	3
2NC	2	2	4	2	5	3	2	2	2.75
3NC	1	2	1	4	5	1	1	2	2.125
4NC	2	2	3	2	5	2	2	2	2.5
5NC	2	2	3	2	5	2	1	2	2.375
6NC	2	2	3	5	5	2	2	2	2.875
7NC	4	2	3	4	5	2	2	2	3
8NC	4	2	3	2	5	1	1	2	2.5
9NC	4	2	1	2	5	2	1	2	2.375
10NC	4	2	3	2	5	2	1	2	2.625
11NC	2	2	3	4	5	2	2	2	2.75
12NC	4	2	3	2	3	2	2	2	2.5
13NC	1	2	1	1	1	1	1	1	1.125
14NC	2	2	4	4	5	2	2	2	2.875
15NC	4	2	3	4	5	2	2	2	3

Table 6-3: The accuracy of participants' interpretations in both 'visual context' and 'no visual context' groups.

	‘Visual context’ descriptive statistics					‘No visual context’ descriptive statistics				
	N	Min	Max	Mean	Std. Deviation	N	Min	Max	Mean	Std. Deviation
Email_2_FR	15	1	5	3.07	1.335	15	1	4	2.80	1.207
Like_3	15	2	5	3.00	.926	15	1	5	2.93	1.223
Like_4_FR	15	1	4	2.27	.704	15	1	2	1.87	.352
Review_4_FR	15	1	5	2.13	.834	15	1	2	1.60	.507
Password_2_FR	15	1	2	1.87	.352	15	2	2	2.00	.000
UserAcc_4_FR	15	2	5	4.27	1.163	15	1	5	2.07	.961
Map_3	15	3	5	3.93	1.033	15	1	5	4.60	1.121
Map_2_BK	15	2	5	3.13	.834	15	1	4	2.60	1.056
Valid N (listwise)	15					15				

Table 6-4: Descriptive statistics of the accuracy scores for the ‘visual context’ group (on the left) and the ‘no visual context’ group (on the right).

RQ3 was addressed by investigating the effects of visual context on participants’ abilities to accurately interpret the meaning of icons. As stated above, the IV for this research question was the visual context (visual context and no-visual context). The DV for this research question was the accuracy of interpretation. The hypothesis and the null hypothesis were:

- **H_3:** Participants who are provided with visual context when interpreting icons will be significantly more accurate at interpreting the icons than participants who are not provided with visual context.
- **H_0:** (null hypothesis) Participants who are provided with visual context when interpreting icons will not be significantly different at accurately interpreting the icons than participants who are not provided with visual context.

To address the hypothesis of RQ3, the participants’ accuracy scores for the eight icons were combined into an overall accuracy score for each participant. The analysis for RQ3 focused on comparing the two groups’ (visual context and no-visual context) abilities to accurately interpret the meaning of the icons.

The initial analysis plan for RQ3 was to conduct an independent samples t-test due to the between-subjects design of the experiment (Deviant, 2011). The independent samples t-test requirements are as follows:

1. The independent samples t-test requires that there are no significant outliers. To test for this the data points were graphed in boxplots and z-scores were calculated. The boxplot (see Appendix-6.6) and z-scores indicated that for the no-visual context group there was one significant outlier (z-score +/- 2.68).
2. The independent samples t-test also requires that the data are normally distributed. To check if the data were normally distributed, a Shapiro-Wilk

normality test was run (this is the preferred test when there are less than 50 participants). The Shapiro-Wilk test for the visual context group was not significant, indicating that the data were normally distributed. The Shapiro-Wilk test for the no-visual context group was significant, *W*(15) = .79, *p* = .003, indicating that the data were not normally distributed, which was also reflected in the histogram (see Appendix-6.7). However, after removal of the outlying participant; the 'no visual context' group passed the Shapiro-Wilk test, *W* (14) = .93, *p* = .306. This participant (ID: 13NC in Table 6-3) was removed in order to conduct the independent samples t-test.

3. Finally, homogeneity of variances was tested using Levene's test for homogeneity of variances. This test was not significant and the assumption of homogeneity of variances was met.

After removing the one outlying participant, the data passed all assumptions necessary to run an independent samples t-test. In addition to the main analysis, the researcher was also interested in seeing if the two groups differed in their ability to accurately interpret the eight icons individually. For this analysis, the accuracy scores were not averaged across the eight icons, which meant that the accuracy scores were on a five-point ordinal scale. Due to the ordinal nature of the dependent variable, the Mann-Whitney test was used for these analyses.

6.3 Results

The independent samples t-test indicated that there was a significant effect of visual context, *t* (27) = 2.51, *p* = .018, 95% CI [.05, .54], with the visual context group having higher accuracy scores (*M* = 2.96, *SD* = .36) than the no-visual context group (*M* = 2.66, *SD* = .27). To further investigate these results, the two groups were then compared on their ability to interpret the eight icons individually. At the one-tailed exact significance level, the Mann-Whitney test revealed a significant difference between the visual context and no-visual context groups for the user account icon (icons 6), *U* =23.00, *p* < .001. The visual context group was more accurate (median = *5*) at interpreting the user account icon (Icon 6) than the no-visual context group (median = 2). The results of the Mann-Whitney test can be seen in Figure 6-5.

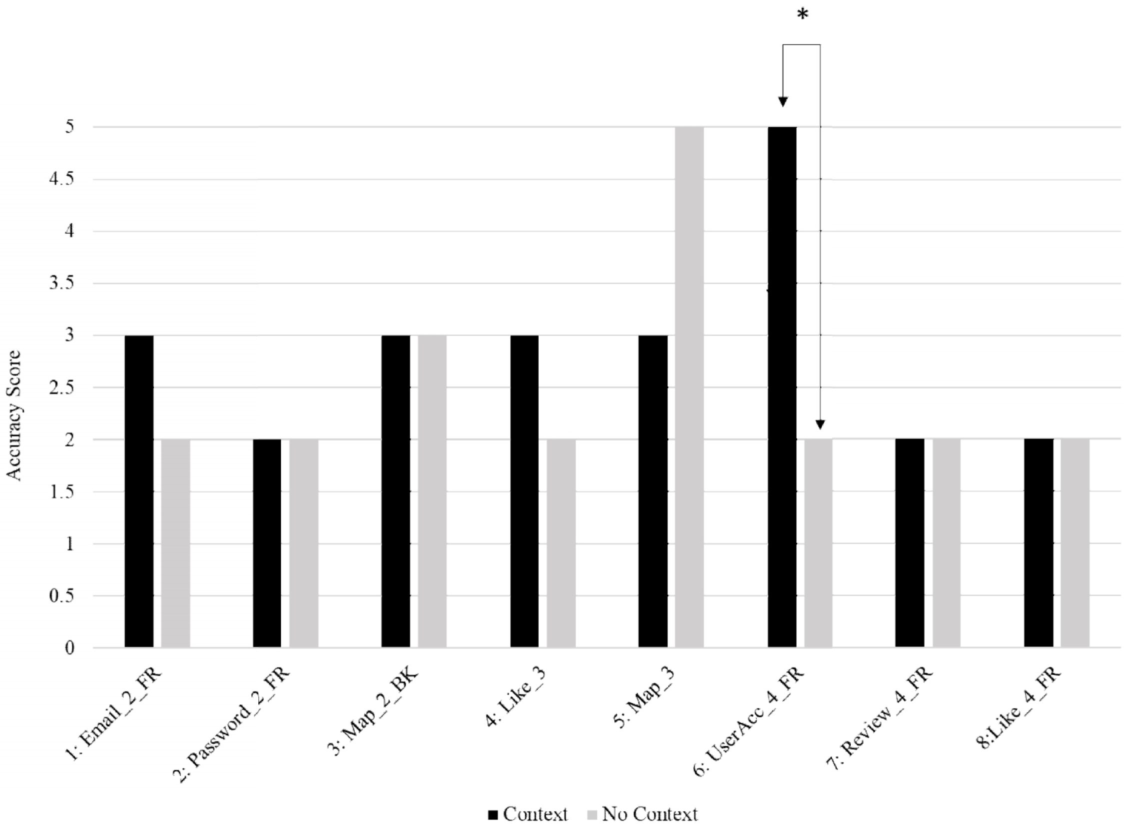


Figure 6-5: Accuracy scores (y-axis) for the visual context group (black) and the no-visual context (grey) group. Asterisks (*) indicate statistically significant results.

6.4 Discussion

Study III asked the question, 'What is the impact of the visual context on participants' interpretations of icons?' To investigate this question, two groups were compared on their ability to accurately interpret eight icons which were selected from Study II. The hypothesis for RQ3 stated that participants who are provided with visual context when interpreting icons will be significantly more accurate at interpreting the icons than participants who are not provided with visual context. The hypothesis of Study III was supported. Participants who were provided with visual context were more accurate at interpreting the meaning of the icons than participants who were not provided visual context. However, further investigation revealed that this result may have been driven by the user account icon.

When the eight icons were analysed individually, the groups were only significantly different at accurately interpreting the user account icon. The visual context group was more accurate at interpreting this icon than the no-visual context group. Referring to Table 6-2, the purpose of this icon (briefcase with a British pound sign) is to add money to the user's account. Here, the meaning of the standing-alone icon was difficult to interpret as the participants may have not experienced this icon previously; therefore, the visual context helped the participants understand the purpose of the icon. In general, the findings of this study were more difficult to interpret than expected. It seems that multiple factors may have influenced the participants' interpretation of the icons, such as their prior experience with similar icons and their use of the visual form of the icon to help them interpret the icon. Hence, it seems that participants do not rely only on the visual context to interpret the meaning of an icon.

Firstly, potential reasons why there was no significant difference in the accuracy of interpretations between the groups for five of the icons are now discussed. Referring to Table 6-2, the purpose of Icon 1 (envelope with a discount tag) in the e-mail family was to ask for the discount code via e-mail, the purpose of Icon 3 (map with location point) in the map family was to show the route between the hotel and the city centre on the map, the purpose of Icon 4 (thumbs-up) in the like family was to show the number of customers who liked the food provided in the hotel, the purpose of Icon 5 (map with silverware) in the map family was to show the nearest restaurants to the hotel on the map, and the purpose of Icon 8 (heart with broom) in the like family was to show the number of customers who liked the cleanliness of the hotel. A first possible explanation is that the two groups ('visual context' and 'no-visual context') were not significantly different at accurately interpreting these icons due to information presented in the icons themselves. For example, for the map_3 icon, which shows a map with silverware, the participants may have recognised the silverware and associated this image with eating. Furthermore, the two groups may not have been significantly different at accurately interpreting these icons due to prior knowledge of the icons. For example, e-mail is often represented by an envelope, hence the participants in both groups may have determined that this icon had something to do with e-mail. Maps are also frequently used as icons; hence the participants were probably able to determine that the map icons had something to do with maps, directions, and locations. The thumbs-up icon is used by users of Facebook when they want to 'like' a post, hence it makes sense that participants in both groups probably determined that this icon had to do with liking. On other websites (for example, dating websites) the heart icon is used when the user wants to 'love' someone, and on Facebook, it is used when a user 'loves' a post. Hence, it makes sense that participants in both groups probably determined that this icon had to do with liking or perhaps loving something. While the icons in this study were not identical to icons with which the participants may have had prior experience, it is possible that the participants did not look at the icons in the experiment closely. Perhaps they quickly recognised the thumbs-up icon and the heart icon and associated them with

Facebook and did not look at the icons for further detail. Furthermore, it could be that participants were influenced both by the information presented in the icons and their prior knowledge.

The groups were also not significantly different at accurately interpreting the password icon (Icon 2). Referring to Table 6-2, the purpose of Icon 2 (lock with an offer tag) in the password family was to show the secret flight offer for today. It is unclear how the icon itself could help the participant determine its purpose, however, the 'visual context' group (*M* = 1.87, *SD* = 0.35) and the 'no-visual context' group (*M* = 2.00, *SD* = 0.00) were both quite inaccurate at interpreting the meaning of the icon. Although the groups were not significantly different in their interpretation of the icon, they were both quite inaccurate at interpreting the meaning of the icon. An accuracy score of 2 indicated erroneous accuracy and an accuracy score of 1 indicated that the participant was incapable of recognising the meaning of the icon. Therefore, in this case, perhaps there was no information in the icon which helped the participants identify the meaning of the icon. The two groups were also not significantly different at accurately interpreting the review icon (Icon 7). Referring to Table 6-2, the purpose of Icon 7 (a staff with a review sign) in the review family was to show customers' reviews of the staff in the hotel. It is unclear how the icon itself could help the participant determine its purpose, however, the 'visual context' group (*M* = 2.13, *SD* = 0.83) and the 'no visual context' group (*M* = 1.60, *SD* = 0.51) were both quite inaccurate at interpreting the meaning of the icon. Based on these results, it seems that the visual context did not benefit in the interpretation of icons 1 and 7.

In this study, participants were placed in one of two groups; either a 'visual context' or 'no visual context' group. Both groups were required to determine the meaning of the eight icons in this experiment. It was hypothesised that the 'visual context' group would be significantly more accurate at interpreting the meaning of the eight icons than the 'no visual context' group. However, the results of this study showed that the 'visual context' group was only significantly more accurate than the 'no visual context' group at interpreting Icon 6. Here, participants in both groups tended to cite that the icon was related to 'money', however, the participants in the 'visual context' group were able to determine that the icon's function was 'to add money to the user's account'. Hence, it seems both groups used visual information about the icon to determine its purpose (e.g., pound sign in the icon), however, when the icon was presented in its visual context, participants were able to fully determine its meaning.

On closer inspection of the participants' responses, it was seen that for some of the icons the participants determined the icons' meanings based on visual information in the icon and/or prior experience with similar icons. For example, for Icon 1 'envelope with a discount tag', participants often interpreted the icon as 'email'. For Icons 3 and 5, the participants often interpreted the icons as 'map'. For icons 8 and 4, the participants tended to rely on information about similar icons they had seen before (e.g., the Facebook 'like' icon, which is a thumbs-up or the Facebook 'love' icon, which is a heart). For Icon 2 'lock with an offer tag', participants in both groups were inaccurate at identifying the icon's meaning. Participants in both groups tended to interpret the meaning of the icon as 'lock', hence it seems they were using visual information about the icon to try to determine its function. However, as the meaning of the icon was 'to show the secret flight offer for today', and 'lock' was an inaccurate interpretation. Furthermore, for Icon 7 'staff with a review sign', it seems that both groups may have used visual information to try to determine the meaning of the icon. Typically, responses were 'staff', 'person', 'baggage', and 'concierge services'. Furthermore, one person in the 'visual context' group and six people in the 'no visual context' group wrote 'I do not know' when trying to interpret the meaning of the icon. Hence, it seems this icon was hard to interpret and that participants tried to use visual information about the icon to interpret its meaning.

A limitation of this study was the small number of participants. Although the sample size of Study III was reasonable, in order to increase the statistical power of Study III it would be beneficial to re-run this experiment with a larger sample size. Another limitation was that the participants were recruited online via Mturk. Therefore, the experimenter did not have complete control over the selection of the participants or the accuracy of their data.

The researcher set-up Mturk so participants could only be in one of the two groups, but not both. However, no control can be obtained over whether the participants were familiar with the icons before participating in the studies. There was no guarantee that someone who participated in Study I or Study II did not participate in Study III. If someone did participate in Study II, then participated in Study III, they would have been more familiar with the icons than the other participants.

Furthermore, although an outside coder coded the responses for Study III, this could have been a limitation in that the out-side coder was not as familiar with the experiment as the researcher. In the future, it may be more beneficial to have both the researcher and an out-side coder code the data, and then calculate inter-rater reliability coefficient between the two coders' codes. Furthermore, regarding coding, Study III used a five-point accuracy scale; this scale was used in order to provide more information about the participants' responses and to understand to what degree their responses could vary. However, in the future, it may be beneficial to use a binary accuracy scale wherein the participants' responses are considered correct or incorrect. This may reduce the variability in the individual judgement of the coders. With a binary scale, the participants' responses could be considered correct if the participant interpreted the meaning of the icon exactly as the designer intended and all other responses could be considered incorrect.

Finally, for seven of the eight icons in this study, the visual context did not seem to help the participants interpret the icons. As was stated before, this may be due to the participants' interpretations of the icons being influenced by the information presented in the icons, their prior knowledge of similar icons, or both. In the future, it may be beneficial to use icons similar to those in this study, but to explicitly ask users if they used visual elements of the icons to help them interpret the icon and if they have prior experience with icons that are visually similar to those in the study. Another option for future studies would be to use only icons that do not contain extra visual information and/or to use icons that the participants do not have prior knowledge about. For example, many people are familiar with the thumbs-up icon and the heart icon due to their usage of Facebook, hence, such icons may not be the best to use in future experiments.

The complexity of the meaning of the icons may have hindered user performance. For example, for Icon 1, the meaning was 'to ask for the discount code via e-mail', in this case understanding that the icon is so specific (e.g., ask for the discount code) may hinder user performance. Perhaps when icons have a complex meaning better visual context needs to be provided in order to help the user interpret the icon.

As a further consideration, the type and amount of visual context were different across the images (where the icons were placed). For example, there were variable amounts of visual elements across the images which led to the images having different amounts of visual information. The varying amount of visual information across images may have played a role in the users' interpretations of the icons. Therefore, in the future, it may be beneficial to control the visual context of the images that the icons were presented in. For example, in the future, it may be beneficial to present the icons on the same image. Using the same image would allow the user to experience the same type and amount of visual context in the image and across icons. Therefore, if the users were using the visual context from the image to interpret the icon, the visual context they would be using would now be the same for all icons.

6.5 Conclusion

Based on the observations, it seems that the visual context has the potential to help participants interpret icons. However, if the icon is visually or semantically like other icons the participant has experienced before, they may have trouble interpreting an icon's meaning regardless of whether they are provided with visual context. Furthermore, it seems that when icons are unfamiliar, participants rely heavily on components of the icon to interpret its meaning, regardless of whether the icon is in its visual context. It is suggested that in future work novel icons which are not visually or semantically like icons that the participants may have interpreted before are used.

Chapter 7 Summary and Conclusions

This chapter presents a summary of the research conducted to meet the three objectives of this PhD. For each study, the key findings, the interpretations of the findings, the implications of the findings, the limitations and future work are discussed. After that, overviews and reflections on the research as a whole from the following perspectives are presented: the study design, the participants, the data collection process and the data analysis. Finally, in a broader overview, the findings are considered from the perspective of Gestalt effects, and the contributions of the research, its limitations, potential future research and closing remarks are presented.

7.1 Objectives

The main research question of this thesis was: *how do participants determine the meaning of similar and dissimilar icons?*

Three objectives were identified to address the main research question. One exploratory study (Study I) and two experimental studies (Study II and Study III) were conducted to meet the three objectives. The three studies asked eight research questions.

7.1.1 Objective 1: To explore the effect of presenting ambiguous icons in different orders on the accuracy of users' interpretations

The first objective of this research was met by conducting Study I. Study I addressed the following research questions:

- *RQ1.1: When viewing ambiguous icons on a limited-size screen, does the order in which icons are presented to participants influence how they interpret those icons?*
- *RQ1.2: What reasons do participants give for their interpretations?*

7.1.1.1 Summary of Key Findings

Study I compared two groups, one who completed a flight booking task first and one who completed a hotel booking task first. Results indicated that the two groups were notably different in their accuracy when interpreting the ambiguous icons in the pre-interaction phase. The flight-first group interpreted the icons accurately more than twice as often as the hotel-first group. Furthermore, the hotel-first group was twice as likely to be incapable of interpreting the icons correctly, as compared to the flight-first group. Additionally, the hotel-first and the flight-first groups were almost equally able to interpret the icons 'moderately'. The groups were not notably different in terms of the number of icons which were interpreted conflictingly or erroneously. In the post-interaction phase, both groups tended to be accurate in their interpretation. When comparing the participants' accuracy pre-interaction and post-interaction, participants in both groups were more accurate at interpreting the icons in the post-interaction phase.

Furthermore, during the post-interaction phase, the participants were asked, 'Do you think this sign adequately represents its intended function on the interface? Why?' The participants' answers were rated by the researcher on a five-point scale from 1: very adequate to 5: very inadequate. An example of adequacy scoring was given in chapter 4. In this example, when a participant in the flight-first group was asked about the adequacy of the heart icon during the flight task, they said, 'No. I don't think so, because this one means I like it; the heart shape means I like something'. However, the heart icon was used during the flight task to mean 'save the flight details to the user's favourite list'. Hence, this participant did not understand the meaning of the icon in the flight-booking scenario and their response was coded as *inadequate*. With regard to the level of adequacy perceived by participants in the post-interaction phase, both groups showed similar patterns, except for the *very adequate* category where the flight-first group was twice as *very adequate* as the hotel-first group. Furthermore, in the pre-interaction phase, the participants' confidence in icon identification was compared between the first time they saw an icon (the first icon in the pair) and the second time they saw an icon (the second icon in the pair). Results from this analysis showed that participants were more confident at identifying an icon's meaning the second time they saw it.

As the analysis was conducted for the second icon in each pair, the results indicated that the hotel-first group cited *order of icons* more often as their reason for interpreting the meaning of the ambiguous icons the way they did. In both groups, the order of icons was cited by the same number of participants for Icon 5 (hand icon) and Icon 6 (box icon). For Icon 4, the order of icons was cited slightly more in the hotel-first group (nine participants) than in the flight-first group (eight participants). It seemed that the order in which icons were presented and the context in which an icon is presented could influence the participants' interpretations. The impact of the visual context in which an icon is presented was further investigated in Study III.

7.1.1.2 Interpretations of Findings

In Study I, when participants were provided with the meaning of an ambiguous icon and then later asked to interpret the meaning of a visually identical icon with a different meaning, their interpretation of the second icon was often not correct. This pattern was not only seen in the accuracy of participants' responses but was also reflected in their verbal explanations of how they interpreted the icons. These results indicate that perhaps the participants utilised short-term memory to remember the function of the icon over time. It is speculated that the participants learned (kept in memory) the meaning of the first icon they were first presented with, then when they were presented with the second icon (which was visually identical to the first icon, but had a different meaning) they utilised the information they had initially learned to interpret the meaning of the second icon. However, since the meaning of the second icon was not the same as the first icon, they were not able to accurately interpret the meaning of the second icon based on their memory of the meaning of the first icon. In all icon pairs, participants recalled information from the previous ambiguous icon, suggesting that recency effects might be at play.

Study I was inspired by the semiotic analysis to evaluate how well the icon designer's intended meaning for an icon and the participant's interpretation of the icons' meanings match. As such, semiotic analysis can be used to investigate the underlying process that occurs when users are asked to interpret icons. In Study I, the semiotic analysis suggested that participants understood the icon designer's meaning of the first icon that they were presented with, but when they were presented with the second icon, they were not able to understand the icon designer's meaning, perhaps due to the influence of the first icon.

This speculation is supported by the reasons that the participants gave for their interpretation of the second icon. For example, when asked why they interpreted the second icon the way they did, one participant stated, 'Mmmm. Mainly because of my previous task'. Furthermore, when the participants saw the icons for the second time, they were more confident about their interpretation. For example, one participant who completed the flight task first and then the hotel task stated, 'When I was in the flight's page it was like something new; new icons; but when I transferred to the hotel's page, I was confident with the icons given'. These results support the idea that the participants tended to believe that the meaning of the second icon would be the same as the first icon.

Furthermore, the participants' responses support the concept of the recency effect. The recency effect suggests that items that have been recently accessed are easier to remember due to their storage in short-term memory (Burgess and Hitch, 2005). Based on the participants' reasons for interpreting the icons and their confidence in their interpretations, it seems that the participants stored information about the first ambiguous icon in their short-term memory, then used this information to interpret the second icon, which was visually identical to the first icon but had a different meaning.

Furthermore, as this thesis utilised a sequence of f research processes, Study I not only provided an answer to RQ1.1 and RQ1.2, but also provided information for the design of the subsequent study (Study II). In Study I, the heart icon pair (Icon 1 and Icon 4), and the box with an upward-pointing arrow icon pair (Icon 3 and Icon 6) were mainly used to answer RQ1.1 and RQ1.2. The icons in the heart and box pairs were visually identical; however, the second icon in each pair (Icon 4 and Icon 6) had a different meaning than the first icon, and the participants tended to be inaccurate at interpreting the meaning of the second icon. However, Icons 2 and 5 (thumb pointing upward and thumb pointing downward) were marginally visually different, and when viewing the second icon in this pair some participants interpreted the second icon 'accurately' or 'conflictingly', indicating that the marginal visual difference in these icons might affect the participants' interpretations of these icons. Although not addressed in the research question of Study I, these findings led to the topic which was addressed in Study II. Hence, in Study II how the degree of visual similarity of icons influenced participants' viewing time and accuracy was investigated.

Furthermore, in Study I, when inductive qualitative coding was applied to the participants' reasons as to why they interpreted the meaning of the icons in the way they did, four main codes and eleven sub-codes emerged. One of the main codes was Icon. This was used when the participants referred to features of the icon or referred to elements that existed in the current/same interface as the reason for their interpretation. The Icon code consisted of the following sub-codes: shape, position, direction, confusion, neighbouring effects, lack of familiarity, and intuitiveness. The second main code was External Context. The External Context code was used when a participant attributed their interpretation of an icon to their previous knowledge or to their previous experience. The code consisted of two sub-codes: unspecific previous experience and specific previous experience. Unspecific previous experience referred to situations where the participant stated that they had used such an icon before, but they did not say when or where. Specific previous experience referred to situations where the participant stated that they had used such an icon before and stated when or where (e.g., WhatsApp or Facebook). The third main code was Order of Icons, which consisted of the sub-codes: previous task/process and previous icon/web page/page. This code was used when the participant attributed their interpretation to the fact that they had seen the icon on the previous page or previous task in the study. The last main code was Multiple Contexts, where participants attributed more than one of the main codes in their responses.

7.1.1.3 Implication of Findings

The findings of Study I suggest that participants used their short-term memory in order to complete the task. These findings support the ‘recency effect’, wherein people recall information that has been recently memorised or accessed more easily than older information (Murphy *et al.*, 2006; Burgess and Hitch, 2005). The findings of Study I indicated variation in the accuracy of participants’ interpretation when they interpreted ambiguous icons in two different orders of presentation. Also, the order of icons was the most frequently cited reason when participants justified their interpretations of ambiguous icons. These findings have relevance to design practice in that icons that have different meanings should be designed to be visually different. For icon designers to create icons that are easily interpretable by the participant, it is important that icons with different meanings do not look alike. If icons look alike, but have different meanings, the user may be easily confused about the purpose of the two icons.

7.1.1.4 Limitations and Future Work

Regarding Study I, two icon pairs were used where the icons in each pair were visually identical, but had a different meaning; however, the icons in the third pair of icons (Icon 2: thumb pointing upward and Icon 5: thumb pointing downward) were marginally visually different (and the icons also had different meanings). As Study I was an exploratory study, this third pair of icons was included in order to provide information for Study II. In the future, it may be beneficial to conduct a study with only icon pairs that are visually identical or icon pairs that are marginally visually different. By combining these two visually similar types in one study, the study had less data for each type and the icons could not be combined for analysis. Moreover, as Study I was an exploratory study, only descriptive statistics were used in reporting the quantitative data. However, it might be useful in the future to conduct the study as an experimental study and run a statistical analysis on the data.

7.1.2 Objective 2: To identify the impact of icons' visual similarity on participants' speed of recognition and the accuracy of their interpretations

The second objective of this research was met by conducting Study II. Study II asked the following research questions:

- *RQ2.1: Will participants who are presented with visually similar icons in a 'forward' order recognise the icons at a different speed than participants who are presented with visually similar icons in a 'backward' order?*
- *RQ2.2: What is the impact of the degree of visual similarity between icons on the speed with which a participant recognises the icons?*
- *RQ2.3: Will participants who are presented with visually similar icons in a 'forward' order interpret the icons with different accuracy than participants who are presented with visually similar icons in a 'backward' order?*
- *RQ2.4: What is the impact of the degree of visual similarity between icons on the accuracy with which a participant interprets the icons?*
- *RQ2.5: Is there a correlation between speed of recognition and accuracy of interpretation for each icon type (the four icon types)?*

7.1.2.1 Summary of Key Findings

Study II investigated the impact of two independent variables on the speed of recognition and accuracy of interpretation. The independent variables were two presentation orders ('forward' and 'backward'), and four icon types (identical, slightly different, mostly different and dissimilar). The experiment was conducted using two tasks types (flight booking and hotel booking). The two task types were counterbalanced while retaining the same presentation order for the icons. This resulted in four experimental groups. For analysis, the data from the task types were combined. The presentation groups were compared on their speed of recognition (RQ2.1) and accuracy of interpretation (RQ2.3) for the four icon types. There was no 'main' effect of presentation order on the participants' speed of recognition or accuracy; however, there was a 'main' effect of icon type. RQ2.2 investigated the effect of icon type on the speed of recognition, and it was found that there was a trend for the participants' speed of recognition to decrease as the dissimilarity of the icons increased. RQ2.4 investigated the effect of icon type on the participants' accuracy, and it was found that there was a trend for the participants' accuracy of icon interpretation to decrease as the dissimilarity of the icons increased. RQ2.5 investigated the possible correlation between speed of recognition and accuracy for the four icon types; however, no significant correlations were found.

7.1.2.2 Interpretation of Findings

Study II investigated the impact of two independent variables on the speed of recognition and accuracy of interpretation. The independent variables were: presentation orders (two presentation orders) and icon types (four icon types). The experiment was conducted using two task types (flight booking and hotel booking) which were counterbalanced while repeating each presentation order across the task types. The results indicated that there was a trend for the participants' speed of recognition to decrease as the dissimilarity of the icons increased. Also, the results indicated that there was a trend for the participants' accuracy of icon interpretation to decrease as the dissimilarity of the icons increased. The length of time it takes a participant to recognise and interpret an icon has been linked to cognitive processes and attention, with slower speed of recognition indicating the use of more cognitive processes and attention. Finally, there was no significant correlation between the speed of recognition and the accuracy for the four icon types. It may be the case that no significant results were found for these correlations, because there was not a strong monotonic relationship between the variables. The main findings of Study II were: (1) as icon dissimilarity increases so does viewing time, (2) identical icons lead to higher accuracy, while in general dissimilar icons lead to lower accuracy, and (3) icon visual similarity may affect cognitive processes and attention.

While Study II provided answers to RQ2.1, RQ2.2, RQ2.3, RQ2.4, and RQ2.5 it also informed the experiment for Study III. In Study II, like in Study I, data were collected about the participants' reasons for believing an icon would behave in a certain way. Inductive coding of this data resulted in nine categories of reasoning:

- Shape.
- Specific previous website (external to the study).
- Unspecific previous website (external to the study).
- Internal to the study.
- Text inside the icon.
- Internal context within the icon.
- Task-related.

- Real-world analogy.
- Icon location.

Information from this analysis showed (like Study I), that when repetitive icons are used, the participants' interpretation of the second or later icon is influenced. That is, if a participant sees an icon, then later sees a visually similar icon, their interpretation of the second icon will likely be influenced by their experience with the first icon. While neither Study I nor Study II explicitly address context, the results of both studies seemed to imply that the visual context in which the icons were presented influenced the participant's interpretation of the icons. This observation leads to further investigation of the icon's visual context in Study III.

7.1.2.3 Implication of Findings

In Study II, there was a trend for participants' speed of recognition to decrease as the similarity between a known 'target' icon and the similar/dissimilar icons decreased. Furthermore, there was a trend for participants' accuracy of interpretation to decrease as the similarity between a known 'target' icon and the test icon decreased. Viewing times (speed of recognition data) were collected with an eye-tracker, and prior research has shown that eye-tracking data equates to attention (Jabeen, 2010) and cognitive processing (Griffin *et al.*, 2004). Hence, it seems that as the icons' similarity decreased, the participants' speed of recognition decreased, indicating that the participants had to use the cognitive process of attention more to interpret the visually dissimilar icons than the visually similar icons. These results were supported by viewing time data which showed that it took participants more time to decide about the meaning of the more visually dissimilar icons. These results support the work of prior researchers who have shown that simple icons elicit quicker responses than complex icons (McDougall, Bruijn, and Curry, 2000; Schröder and Ziefle, 2008). The results of Study II also support prior research which has shown that icons which are more concrete and have a smaller articulated distance elicit quicker responses than less concrete icons with a larger articulated distance (Rogers, 1986; Blankenberger and Hahn, 1991; Cheng and Patterson, 2007).

Study II indicated that it takes participants longer to interpret dissimilar or unfamiliar icons. Hence, in practice, it may be beneficial for icon designers to use icons that are familiar to users; for example, using an envelope icon to indicate e-mail. Furthermore, it may be beneficial for icon designers to standardise the design of icons; for example, having icon designers agree that an envelope icon will always indicate email. However, on the other hand, standardising the design of icons might not be an ideal solution due to the use of multiple applications and web interfaces on different mobile devices or platforms. When designers are creating icons the design of the icons needs to fit a particular design related to the user interfaces or the icon's applications. Hence, it may be that particular user interfaces or applications would not allow a designer to use standardised icons. For example, on a very small device it could be that that the limited screen space does not allow for the detail that icons on larger screen devices can contain. Furthermore, Study II has implications for when a designer needs to design an icon that is like a known 'target' icon. Study II showed that when the dissimilarity of an icon to a known 'target' icon increased, the users' cognitive load increased (i.e. increased attention). Hence, when a designer needs to design an icon that is like a known 'target' icon, the designed icon should be as like the known 'target' icon as possible in order to reduce the cognitive load of users.

7.1.2.4 Limitations and Future Work

Study II was conducted using two task types (hotel booking and flight booking). The design of the study had been counterbalanced to prevent any confounding effects during experimentation. During the analysis the data from the two task types were combined; however, these two task types consisted of six different icon families, whose icons were different. Therefore, a limitation with this study is that any possible effects of the tasks were ignored. Furthermore, during the analysis of RQ2.2 and RQ2.4, the data from the 'forward' and 'backward' presentation groups were combined; however, each order had a different, slightly different and dissimilar icon type. Hence, in the future, it would be beneficial to design a study where only one task order is used or where the two task orders are analysed separately. Furthermore, two presentation orders should consist of the same icons.

Additionally, in Study II, the visual similarity of the icons was determined. There is no single standard method for assessing the visual similarity between icons. In Study II the researcher created his own method of assessing visual similarity. This method consisted of assessing the number of visual fractions (characteristics) that were shared between a known 'target' icon and a test (similar/dissimilar) icon. This is a novel method for assessing the visual similarity between icons; hence, in the future, it may be beneficial to apply this method repeatedly in order to test its validity and reliability. Furthermore, in the visual similarity method proposed in Study II the visual similarity of icons was judged based on whether two icons were different by one visual fraction, two visual fractions or more, and no shared visual fraction. These variations produced three icons which were similar and dissimilar to a known ‘target’ icon. In future studies, it may be possible to consider icons that are similar and dissimilar to a known ‘target’ icon.

It had been observed in Study II that participants were unexpectedly accurate on some of the slightly different, mostly different, and dissimilar icons (for example the Like_3 and Map_3 icons). The reason for that might be because common shapes were used in the design of those icons which participants became familiar with. In the future, it might useful to design icons using shapes that are unlikely to be familiar to participants.

The collected and analysed reasons in Study II as qualitative data could lead us to further discussion in future studies. It seems that users’ performance might be affected by multiple factors (see Table 5-5 in Chapter 5). For instance, the undertaken task might play a role in affecting users’ performance. It would be useful in future studies to investigate the impact of factors such as the undertaken task or the icon location on users’ performance.

Reflecting on Study II, another thing to consider is that the study utilised eye-tracking and verbal data. The eye-tracker uses infrared illuminators and any movement during eye-tracking could ruin or influence the data; hence, it was possible to lose or destroy the eye-tracker data if both visual and verbal data were collected simultaneously. The possibility of losing or destroying the eye-tracking data when verbal response data and eye-tracking data are simultaneously collected led to the segregation of eye-tracking and verbal data. Hence, it may be that the eye-tracking data and the verbal responses do not represent the same processes. For example, perhaps if the participants had given their verbal response during eye-tracking their response would have been different than giving it after eye-tracking.

Using the eye-tracker was fruitful to this research. With the eye-tracker, it was possible to know where participants were looking on a user interface and how long they spent looking at specific AOIs. Indeed, useful data were collected using the eye-tracker. However, dealing with the eye-tracker was a time-consuming and sensitive process. It started with calibrating the participants’ eyes with the eye-tracker. Next, conducting the experiment while the eye-tracker was running. After that, segmenting the recorded videos by time-stamping the participants’ viewing times, starting from the time they looked at the user interface until they indicated they had finished by saying ‘done’. Then, drawing the AOIs accurately on icons to extract the participants’ viewing times.

Finally, calculating the participants' viewing times on each icon. In addition, all the processes required attention to details.

Finally, Study II was conducted in the context of sequential web-pages; however, it would be useful to know how the method used in this research would perform in different contexts such as visual search tasks (Alexander and Zelinsky, 2011; Hout *et al.*, 2016).

7.1.3 Objective 3: To explore the impact of the visual context in which an icon appears on the accuracy of users' interpretations

The third objective of this research was met by conducting Study III, which asked the following question:

- *RQ3: What is the impact of visual context on participants' interpretations of the icons?*

7.1.3.1 Summary of Key Findings

Study III compared a group who viewed eight icons in 'visual context' and a group that received 'no visual context' for the same eight icons. Participants who were provided with the visual context were more accurate at interpreting the meaning of the icons than participants who were not provided with the visual context. When the eight icons were analysed individually, a Mann-Whitney test with a one-tailed exact significance level revealed that the groups were only significantly different at accurately interpreting one icon. The 'visual context' group was more accurate at interpreting this icon (the user account icon) than the 'no visual context' group.

7.1.3.2 Interpretation of Findings

In Study III, eight icons from Study II were used; these icons were selected because the participants in Study II were unexpectedly accurate at interpreting these icons. In Study III, overall, the 'visual context' group was more accurate at interpreting the meaning of the icons than the 'no visual context' group. However, further investigation revealed that this effect may have been driven by one icon. When both groups were compared on their ability to accurately interpret the eight icons individually, the groups were only significantly different in their ability to accurately interpret the user account icon (Icon 6). The intended purpose of Icon 6 (user account icon), which looked like a brief case with British pound signs, was to add money to the user's account. Both groups tended to state that the icon was related to 'money', however, the participants in the 'visual context' group were more accurate in their interpretation, for example stating that the icon's purpose was 'to add money to the user's account'. Hence, it seems both groups used visual information about the icon to determine its purpose (e.g., pound sign in the icon), however, when the icon was presented in its visual context, participants were able to fully determine its meaning.

When considering the other icons used in the experiment it is thought that the groups were not significantly different at interpreting the icons, because the icon itself provided enough information for interpretation and/or the participants had prior knowledge of the icons. Icon 1 (envelope) was intended to be used to ask for a discount code via e-mail, Icon 3 (map with location point) was intended to show the route between the hotel and the city centre on the map, Icon 4 (thumbs-up) was intended to show the number of customers who liked the food provided in the hotel, Icon 5 (map with silverware) was intended to show the restaurants nearest to the hotel on the map, and Icon 8 (heart with broom) was intended to show the number of customers who liked the cleanliness of the hotel. Participants were likely to have seen an envelope icon representing e-mail before, map icons are often used for location information, the thumbs-up is used by Facebook to indicate 'like', and the heart icon is used by Facebook to indicate 'love'. Furthermore, silverware would be indicative of eating.

For Icon 2 (lock with a percentage tag), which was intended to show the secret flight offer for today, it is thought that the two groups were not significantly different at interpreting the icon because they were both so inaccurate at interpreting the icon. Hence, it seems that this icon's visual representation did not provide enough information to help the participants to understand the meaning of the icon, even when they were provided with visual context.

The results of Study III suggest that the visual context has the potential to help participants interpret icons. However, the participant may have prior experience with icons which are visually similar or semantically like the icons being tested. This interference effect can occur both when the participant is provided with the visual context and when the participant is not provided with the visual context for the icon. Furthermore, it seems that when an icon is unfamiliar, participants rely heavily on components of the icon to interpret its meaning. This method of interpreting unfamiliar icons can occur both when the participant is provided with the visual context and when the participant is not provided with the visual context for the icon. These observations suggest that participants used visual components of the icons, their prior knowledge of similar icons, or both in order to interpret the icons, regardless of whether they were provided with the visual context or not.

7.1.3.3 Implication of Findings

Study III showed that providing the visual context with icons has the potential to help participants interpret the meaning of icons, in comparison to participants who are not provided with the visual context. However, in Study III it was only found that the ‘visual context’ group was more accurate than the ‘visual context’ group at interpreting one of the eight icons which were used in the study. The observation that the visual context helped participants interpret this icon supports work by Huang and Bias (2012), who showed that when participants are not familiar with icons and are not given any context, they are inaccurate and inefficient at recognising icons. Study III adds to this work by showing that if icons are too familiar and/or have too much visual information, the participants who are provided with the visual context are no more accurate than those who are not provided with the visual context. Hence, if icons are too familiar, participants may rely on prior representations of the icons to interpret their meaning (e.g., when seeing a thumbs-up, if it means ‘like’, as it does on Facebook).

7.1.3.4 Limitations and Future Work

Regarding Study III, the ‘visual context’ group was significantly better at interpreting the meaning of the icons than the ‘no visual context’ group; however, this result seemed to be driven by one icon. The ‘visual context’ and ‘no visual context’ groups were not statistically significantly different at interpreting the meaning of most of the icons, which may be due to their familiarity with the icons and/or the visual information in the icons. Familiarity refers to situations wherein the icons looked like icons the users had encountered before (e.g. on WhatsApp or Facebook). Visual information refers to pieces of the icon which the users used to interpret the icon; for example, for Icon 7 (which looked like a hotel clerk), the participants responded that the icon represented ‘staff’, ‘person’, ‘baggage’, and ‘concierge services’. However, on the other hand, for one icon (Icon 2: lock with a percentage tag) it seems that even when provided with the visual context, the icon was so hard to interpret that the ‘visual context’ group was no better at interpreting its meaning than the ‘no visual context’ group.

Based on these results, it seems that even when provided with the visual context, participants still might use visual features of the icon, their prior knowledge of similar/familiar icons, or both to interpret the meaning of an icon. One recommendation for future research is that icons like, or the same as, those used in Study III are investigated again; however, this time participants would be explicitly asked if they used visual elements of the icons or their prior experience with icons to help them interpret the icons. Furthermore, in the future, it may be beneficial to conduct a study with icons that do not contain extra visual information and/or are unfamiliar to the participants.

Another consideration is that the icons were viewed in different images, and since these images were different, they each had a different type and amount of visual context. For example, the images had varying amounts of visual elements which influenced the type and amount of visual context. The participants' interpretations of the icons may have been influenced by the different images that the icons were presented in. In future work, it may be beneficial to use the same image in which to present each icon. If the same image is used with each icon any influence from the visual context of the image may be controlled across the icons.

Another limitation of this study was the small number of participants. Although the sample size of Study III was reasonable, in order to increase the statistical power of Study III it would be beneficial to re-run this experiment with a larger sample size.

Finally, Study III was conducted without considering the context of the tasks in which the icons appeared. In future work, it may be beneficial to investigate the impact of the context of tasks as well as the visual context to know whether the context of tasks would affect users' interpretation.

7.2 Overview and Reflections

Having considered each study individually, this section now turns to a consideration of the studies. It provides an overview and reflection on the design of the studies, the participants, the data collection process and the data analysis.

7.2.1 Design

Each of the three studies had its own specific design. However, the general research process for Study I and Study II was:

1. Identify (Study I and Study II) and design similar icons (Study II).
2. Develop a prototype website.
3. Recruit participants.
4. Collect data.
5. Undertake inductive and deductive coding.

6. Analyse data.
7. Report and discuss findings.

The differences in Study II were that after developing the website, but before recruiting the participants, expert tests were conducted to evaluate the visual and semantic aspects of the similarity families of icons. Furthermore, in Study II eye-tracking was used and the data had to be extracted from the eye-tracker before they could be analysed. Study III followed a similar design; however, in Study III, instead of identifying similar icons and developing a website, icons were selected from Study II and were presented to participants online via Qualtrics. Study design figures are presented for Study I in Chapter 4, for Study II in Chapter 5, and for Study III in Chapter 6.

The design and conduct of Studies I and III were relatively straightforward as they investigated limited factors, and the eye-tracker was not used. Nonetheless, it was time-consuming to design the icons and web-site, run the sessions and code the data. The design of Study II was complex due to the number and the variety of processes that were involved. Issues considered during the design of Study II included identifying the independent and dependent variables, counterbalancing the order of tasks to prevent any confounding effects, identifying and designing the icons and the visual similarity families of the icons, identifying the meanings of the similar icons in each similarity family and controlling the semantic similarity between the designed icons. An evaluation process was conducted on the designed icons, where experts rated the visual and semantic aspects of the similarity families of icons. During the experiment, an eye-tracking system was used, and the data had to be extracted from the eye-tracker before they could be analysed; this analysis required meticulous attention to detail. In the future, a study such as this could be simplified by dividing it into two separate studies: each study would investigate the impact of one independent variable. Thus, the statistical analysis would be more straightforward.

A Tobii X-60 eye-tracker was used in Study II. This is an old generation eye-tracking system, so the researcher encountered some difficulties in setting up the experiment and during the calibration process. It would be useful to use a new generation eye-tracking system which has advanced technology with enhanced user experience. The new Tobii eye-tracking platform (IS5) has improved and the calibration process has become more user-friendly and easier to do. In the new platform, compared to X-60, setting up the eye-tracking does not need physical measurements for the eye-tracker angle, the distance between the eye-tracker and a participant or the distance between the stand-alone eye-tracker and the monitor. The eye-tracking kits can be easily integrated into laptops, desktops and other devices (Tobii Technology, 2019).

7.2.2 Participants

Different participants were recruited in the three studies. However, the overall criteria for inclusion in the three studies were that the participant was at least 18 years old, had experience with technology and used smartphones frequently, and had normal or corrected to normal vision. Exclusion criteria were that the participant was younger than 18 years old, had little to no experience with technology and smartphones, and did not have normal or corrected to normal vision. For Study I and Study II the participants were recruited to attend the experiment session at a laboratory at City, University of London. For Study III, the participants were recruited online via Amazon Mechanical Turk (Mturk). For all the three studies, convenience sampling was used. Convenience sampling is a non-probability sampling method which consists of recruiting people who are easy to reach to participate in the experiment (Etikan, 2016). Since convenience sampling is a non-probability sampling method, there is no guarantee that every element of the population will be included in the sample. For Study I, a convenience sample of twenty participants was collected and divided evenly into two groups. For Study II, a convenience sample of twenty-four participants was collected and divided evenly into four groups. For Study III, a convenience sample of 30 participants was recruited via Amazon Mechanical Turk (Mturk). The participants were divided evenly into two groups and asked to complete the study via Qualtrics. It would be useful in future studies to use different sampling method such as stratified sampling, which focuses on a variation reduction technique (Tocher, 1963). In stratified sampling, a sample is divided into homogenous subgroups and selected based on a design within each stratum.

Another point about the samples is that most of the participants recruited in this research were students at City, University of London. It would be useful to recruit other people who are for example, not aware of the technology, elderly people or people who are from other cultures, to see how this would affect the results of this research. As suggested previously, if one were to use stratified sampling, a sample with different people of different characteristics could be divided into subgroups based on some pre-defined characteristics.

While the research studies in this thesis have used relatively small sample sizes, other researchers have also conducted different studies with such sample sizes. For example, Freeman *et al.* (2009) recruited twenty-two participants, who were evenly divided into two groups, to participate in a comparative study. The study compared the benefits of using shadow guides-based leaning and video-based instructions learning. Shrestha (2007) conducted a comparative study, comparing user effectiveness, efficiency, and satisfaction between a desktop browser and a mobile browser using 12 participants. De Angeli *et al.* (2006) recruited 28 participants, evenly divided into two groups, to participate in a usability evaluation study. Furthermore, Wiklund *et al.* (2016), suggests that a minimum of 15-25 participants is a 'good working number' for the total sample size. In studies similar to the research conducted in this thesis, Isherwood *et al.* (2007) conducted a study using 30 participants to investigate the impact of icon characteristics such as semantic distance, concreteness, familiarity, and visual complexity, on speed and accuracy of icon identifications. Alexander and Zelinsky (2011) conducted three experiments to investigate the impact of the visual similarity between items on search guidance. In one of these experiments (Experiment 2), they used 24 participants to show the impact of visual similarity between items using distractors in a visual search task.

7.2.3 Data Collection

In Study I, qualitative data were collected via semi-structured interviews and then coded into quantitative data. The researcher asked the participant a series of nine questions before and after they interacted with an icon by clicking on it. These questions were repeated for each icon. The participants were required to answer these questions aloud, verbally. The experiment was audio-recorded, and the participants' responses were later transcribed. Conducting the interviews and transcribing the audio recordings of the interviews were time-consuming processes. During the interviewing process, it was difficult to explain to a few of the participants the exact meaning of some of the questions, such as the question: 'Do you think this icon adequately represents its intended function on the interface?'. Therefore, the researcher was not able to code some responses. It would be useful, in future, to have a brief induction or video-based instructions to explain to participants the interview in detail before starting.

In Study II eye-tracking data were collected from the eye-tracker and qualitative data were collected via questions that the participants had to answer aloud, verbally, using the think-aloud method (reflections on using the eye-tracker in Study II were discussed in section 7.1.2.4). These qualitative data were then coded into quantitative data. Eye-tracking data was collected from the participants the entire time they were involved in the experiment. During the experiment, the participants were presented with icons. The participants were instructed that initially they should view the icons in silence and think about what the icons would do if they clicked on them. After thinking silently about the function of the icons, the participants were instructed to say out loud, 'I am done'. After this the researcher asked the participant two questions:

1. What do you think will happen if you click on each icon?
2. Why do you think that will happen?

After answering these questions, the participants were instructed to click on the icons. After doing this, the participants were instructed to tell the researcher what happened when they clicked on the icons.

Collecting the data and following the protocol while the eye-tracker was running required attention to detail and to follow the steps carefully. Anything could go wrong at any point during the experiment, such as if a participant changed the way she or he was sitting in front of the eye-tracker while undertaking the viewing in the silent phase. This movement could result in the participant moving out of range, so their eyes became untraceable. Thus, the participant eyes' movements could not be collected. The researcher had to pay attention to such details to prevent this from happening.

In Study III, qualitative data were collected via online forms using a text interface and then coded into quantitative data. Study III was conducted online with Qualtrics, the participants viewed eight icons individually via Qualtrics pages; however, they did not click on the icons. On each icon display page, the participants were presented with this text, 'What would be the meaning/function of this icon? Please describe below:'. The participants were required to type their answers into the text box provided. Online recruitment offers an easy way to obtain a larger sample of participants. In Study III, the participants were recruited online via Mturk. However, online recruitment is potentially risky as the researcher does not have complete control over the selection of the participants or the accuracy of their data. However, in this case, the researcher was happy with the quality of the data that were collected online; except for one participant whose data appeared to be inaccurate.

7.2.4 Data Analysis

During Studies I and II the participants were audio-recorded while they verbally gave their answers to questions about the icons asked by the researcher. These audio recordings were transcribed and then coded. Qualitative and quantitative coding were used to aid the researcher in understanding the perspectives of the participants and also to create quantitative data that the researcher could analyse. Studies I and II entailed both deductive and inductive coding. Deductive coding was used for the questions where the participants were asked *what* they thought an icon would do/ what it did do (i.e., 'what' questions). These questions revealed the accuracy of the participants' interpretations of the meanings of the icons. Inductive coding was used for the questions wherein the participants were asked *why* they thought an icon would do what they expected it to (i.e., 'why' questions). These questions revealed the reasoning behind the participants' interpretations of the meanings of the icons. Study III used only deductive coding.

The coding terminology used for deductive coding was adopted from Islam and Bouwan (2016), who categorized participants' understanding of icons into five categories: accurate, moderate, conflicting, erroneous, and incapable. Coding was conducted both manually and using a computer-assisted qualitative data analysis software. For qualitative data analysis, NVivo 11 was used. NVivo is a computer-assisted qualitative data analysis software. NVivo 11 was used to manage the data and aid in the analysis process in Study I. The software was used to query keywords for comparison with manually coded categories and themes. However, the data were first coded by the researcher before being submitted to NVivo 11; as such, NVivo 11 was not the primary coding source and was only used to solidify data analysis. Also, Microsoft Word was used for coding qualitative data analysis throughout this research. Dealing with qualitative data is time-consuming as the researcher had to go through the transcribed texts, extract the data from the transcripts and organise the texts for coding. Coding qualitative data by a human could be biased. Although the used coding scheme was proposed and tested previously by other researchers (Islam and Bouwan, 2016), the researcher went through the data and coded them several times, and each time there was a slight difference in the way the data were coded. Thus, future work could involve a second coder to reduce potential bias (for more details, see section 7.4). Regarding the used five-point accuracy scale; this scale was used in order to provide more information about the participants' responses and to understand to what degree their responses could vary.

The quantitative data analysis was conducted using Statistical Package for Social Sciences (SPSS) software. In order to analyse the data for RQ2.1 and RQ2.3, the Aligned Rank Transform tool (AR Tool) was used before running the statistical test using SPSS. Preparing the data in a certain format into character-delimited (CSV) files to be used as an input for AR Tool was a time-consuming process. After that, the output CSV files from AR Tool were imported to SPSS to run the relevant statistical tests.

Obtaining non-normally distributed data in Study II made the statistical analysis more complicated. It would be useful to investigate this further to determine whether there was a cause for this, in order to avoid it in future studies.

7.3 A Broader Overview

Finally, in a broader overview, the findings of this research are considered from the perspective of the Gestalt principles, and the research contributions, limitations and future work, and closing remarks are presented.

7.3.1 Gestalt Effects

The Gestalt principles were discussed in Chapter 3 and are relevant for interpreting the findings of the three studies. The fact that the human brain innately tries to find patterns, even in items that are not related, shows that the brain places significance on organised structures rather than individual items. As such, the human brain seems to prefer emergent, holistic and contextual perspectives (Soegaard, 2010).

In the experiments conducted in this thesis, it was often observed through the verbal responses of the participants that they interpreted the icons based on visual features within the icons. These observations may indicate that the participants were pre-attentively applying the Gestalt principles to the icons in order to create patterns and organization within the icons, even if none existed. The application of Gestalt principles may have led the participants to see similar patterns in icons, resulting in the participants thinking that these icons were related, when they were not.

Furthermore, as the experiments were conducted on prototype websites, rather than on icons in isolation, it may be that the participants unconsciously applied the Gestalt principles to both the icon they were viewing and other items on the webpage. For example, in Study III the user account icon was near the Add button, which may have resulted in the participants pre-attentively applying the proximity principle. The proximity principle states that items that are close to one another appear to form groups. Hence, the participants may have interpreted the user account icon and the Add button as a group, which may have influenced their interpretation of the user account icon.

7.3.2 Contributions

The primary focus of this thesis was to investigate how participants interpret the meanings of visually similar icons. This thesis has contributed to the investigation of this topic in three ways. Firstly, the research for this thesis has shown that when ambiguous icons are viewed in different orders (icon presentation order), there is an impact on participants' accuracy of interpretations of icons. The findings of Study I suggest that participants used their short-term memory in order to complete the task. These findings support the 'recency effect', whereby people recall information that has been recently memorised or accessed more easily than older information (Murphy *et al.*, 2006; Burgess and Hitch, 2005). However, Study I was conducted in the context of sequential web-pages navigation while the participants were viewing ambiguous icons. This context is important nowadays, because when users are navigating through web-pages they often do so sequentially; therefore, the icons presented on prior pages may influence how the user interprets icons on later pages.

Secondly, the research has revealed how the visual similarity of icons affects participants' speed of recognition and the accuracy of their interpretations. Previous research tended to focus on three icon characteristics: visual complexity, concreteness, and semantic distance (McDougall *et al.*, 2000). It has been indicated in previous studies that simple icons elicit quicker responses than complex icons (McDougall *et al.* 2000; Schröder and Ziefle, 2008). Furthermore, it has been shown that icons that are more concrete and have a smaller articulated distance require shorter reaction times and less cognitive processing than less concrete icons with a larger articulated distance (Rogers, 1986; Blankenberger and Hahn, 1991; Cheng and Patterson, 2007). The research of this thesis has gone beyond these characteristics and investigated the impact of visual similarity between icons on users' performance.

More specifically, as the dissimilarity between an icon and a known 'target' icon increases, participants' speed of recognition decreases. Furthermore, participants were more accurate at interpreting icons that were like known 'target' icons than icons that were dissimilar to known 'target' icons. Taken together, these results may suggest that as icons become less like known 'target' icons, participants need to utilise more cognitive processes, such as attention, in order to interpret the meaning of icons. These findings have real-world implications in that they show that if icon designers want users to quickly recognise and understand icons, then the icons should be familiar to the user.

Thirdly, this research utilised a novel method for determining the visual similarity between icons. There was no standard method for determining the visual similarity between icons. In previous research, it was unclear how to manipulate visual similarity in the range between dissimilar and target icons. Moreover, it was not clear to what extent the visual similarity between icons would affect users' performance.

In this research, the visual similarity of icons was determined based on the number of visual fractions (shapes) shared between the icons. In previous research, the multidimensional scaling method (MDS) was proposed to quantify the visual similarity between items in visual search tasks (Hout *et al.*, 2016). During a visual search task, users were asked to find a target item that was embedded within other non-target items. MDS is a statistical tool that was proposed to know the dimensions by which users perceive an item to be like or unlike another item (similarity). In contrast, the approach in this thesis quantifies the similarity based on the visual characteristics (shapes) of items. In Hout *et al.*'s (2016) work, all the stimuli were selected before using the MDS tool based on participants' subjective ratings. Participants were shown two items at a time and used a Likert scale to rate how similar they were to one another. MDS used participants' ratings to quantify the similarity between the items. Finally, MDS produced a similarity map which showed the similarity between the items. The approach in this thesis uses 'active' similarity value, as the similarity between icons was designed and manipulated before the experiment, whereas in Hout *et al.* (2016), they used 'passive' similarity value as the similarity between items was rated during the experiment and after they were selected.

Alexander and Zelinsky (2011) uniquely proposed an approach to quantify the similarity between items in visual search. They collected visual similarity rankings for two target categories. They had two main categories: 'teddy bears' and 'butterflies. Participants were shown five objects at a time and asked to rank order the objects according to their similarity to the main categories. In the visual search task, three distractors (non-target) categories were displayed: 'low', 'medium' or 'high' similarity to the selected target category. Like MDS in Hout *et al.* (2016), this approach used 'passive' similarity values.

The visual similarity approach in this thesis would allow researchers to design similar icons in the range between a target (similar) icon and a dissimilar icon. Only two icons in this range were designed in this research. This could be taken further by designing more visually similar icons within that range. Furthermore, the method was used on icons, but this could be extended to other images that have visual characteristics and shapes such as images of human faces, animal photos and images of letters.

In this research, the visual similarity criteria were applied by a person (the researcher) in designing similar icons. Determining the visual fraction (shape) in the target icon to be differentiated with icons of other icon types (slightly different, mostly different) is challenging. The subjectivity of human selection for which visual fraction can be used might affect the design of the visual similarity between icons. Moreover, the importance of a fraction (shape) to a whole shape of an icon and its meaning could affect the visual similarity between the target icon and the other icons. For instance, a pin in a map icon could be more important than a dotted line if the meaning of the icon is to show a specific location on a map rather than showing a walking route or a distance between two locations on a map. The visual representation of an icon should represent the meaning of an icon. The visual fraction is part of the whole visual representation of the icon. Indeed, additional similarity criteria could be proposed in future studies to improve the similarity criteria of this research.

The designing of the similar icons could be modelled using an algorithm that takes the target icon as an input and produces three different icons, each of a specific icon type (slightly different, mostly different and dissimilar). Likewise, an algorithm could be investigated to use in the process of assessing the similarity of icons. This might reduce human bias in designing and assessing the similarity between icons. Alternatively, if the human is still involved in the assessment process, more experts could be recruited in the future to obtain more accurate judgments about the visual similarity of icons. In the approach of this thesis, only two experts were asked to rate the visual similarity between icons.

7.3.3 Overall Research Limitations and Future Research

In this research, all three studies were conducted in the context of specific screen sizes that were relatively small compared to laptop and desktop screen sizes. This approach was taken since the research was motivated by the increasing prevalence of devices with smaller screens. The question, therefore, arises as to whether the results reported here would also hold true for larger screen sizes. Several previous studies, such as Jones *et al.* (1999), Shrestha (2007) and Raptis *et al.* (2013), have shown that research studies on large-size screens compared to limited-size screens deliver different results. Schade (2017) suggests that the images displayed on large screens might not be directly suitable for display on small screens and that many problems may arise when using large screen images on small screens. For instance, the clarity and readability of the displayed images might be affected when small screens are used. Moving from one screen size to another might not be straightforward and implies a limitation to the displayed content. Therefore, it cannot be assumed that studies of the impact of icon similarity undertaken on larger screens would deliver the same findings. It would be useful in the future to undertake studies like those reported in this thesis, but with different screen sizes, either smaller or larger than those used here, to determine whether different results would be obtained.

One specific recommendation for future research is to conduct studies with larger sample sizes. Study I had a sample size of twenty with two groups of ten participants. Study II had a sample size of twenty-four participants, who were divided into four groups of six. Study III had a sample of thirty, which was divided into two groups of fifteen participants. As discussed in section 7.2.2, a sample size of fifteen to twenty-five participants is a ‘good working number’ Wiklund *et al.* (2016). However, small sample sizes provide little power to reject the null hypothesis, hence, to improve the statistical power of this research it would be helpful to use larger sample sizes; this may especially be the case for Study II, which only had six participants in each group.

Another recommendation is to use an outside coder or to combine the codes of the researcher and an outside coder. In Study I, the researcher coded the verbal data of the participants, which may have led to bias in the results. In Studies II and III an outside coder coded the data; however, it is possible that this coder was not familiar enough with the experiment to code the data completely accurately. Hence, in the future, it may be beneficial to have an inside and an outside coder to code the data, then to conduct inter-rater reliability analysis on the codes and provide the alpha values in the studies.

7.3.4 Closing Remarks

The study of the visual similarity of icons and its impact on users’ performance in various display sizes is challenging. The rapid evolution of technology, alongside the desire of users to have easier access to devices, leads us to more complicated contextual situations where multiple factors may be at play, affecting users’ performance.

The work in this thesis has advanced understanding of how the impact of visual similarity between icons on users’ performance is influenced by multiple factors. This work could inspire and provide the basis for further study of the impact of visual similarity between icons on users’ performance using various screen sizes. It could also be taken further and move us toward realising the potential impact of visual similarity between icons on users’ performance under multiple contextual factors, for example, a mobile context where users are walking and viewing similar icons on a wearable device, such as an Apple Watch. This would allow the objects that surround the users while they are walking, to place demands on their cognitive resources, thus their performance might be affected (Oulasvirta, 2004).

Chapter 8 Appendices

Appendix-4.1: Participant Interview of Icons in Web Interfaces

(Interview on a Mobile Device)

Date:	Session Time: TO
Type: Semi-structured	Participant Name:
Participant Level:	Recorded: Yes
Mobile:	Interaction Sequence: Flights->Hotels
Confidence Scale: (%0 Not Confident) (%100 Completely Confident)	
Have you ever used holiday booking website in the past? If YES, How many times?	
When was the last time you have used it?	

Interface	*Icon*	*Pre-interaction*		*Post-interaction*	
		Question	*Answer*	*Questions*	*Answers*
Flight Booking-Results	♡	What do you think would happen if you click on this icon?		What happened when you clicked on this icon?	
				Why do you think that happened?	
		Why?		Do you think this icon adequately represents its intended function on the interface? Why? (very adequate...very inadequate)	
		Do you recognise the object represented in this icon? If YES, what is that object?		Do you have any suggestions about how this icon could more clearly represent its intended function?	
		How confident are you about this icon interpretation?			
		Which part of the interface was more helpful to understand this icon's function? (icon only, icon and its context - give percentage % in case of both)			
	☝ 👇	What do you think will happen if you click on this icon?		What happened when you clicked on this icon?	
				Why do you think that happened?	
		Why?		Do you think this icon adequately represents its intended function on the interface? Why?	
		Do you recognise the object represented in this icon? If YES, what is that object?		Do you have any suggestions about how this icon could more clearly represent its intended function?	
		How confident are you about this icon interpretation?			
		Which part of the interface was more helpful to understand this icon's function? (icon only, icon and its context - give percentage % in case of both)			

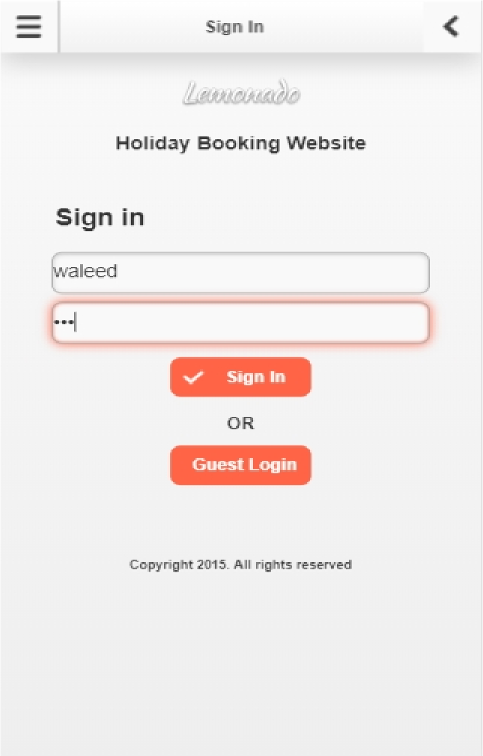
	⍐	What do you think will happen if you click on this icon?		What happened when you clicked on this icon?	
				Why do you think that happened?	
		Why?		Do you think this icon adequately represents its intended function on the interface? Why?	
		Do you recognize the object represented in this icon? If YES, what is that object?		Do you have any suggestions about how this icon could more clearly represent its intended function?	
		How confident are you about this icon interpretation?			
		Which part of Interface was more helpful to understand this icon's function? (icon only, icon and its context - give percentage % in case of both)			
Hotel Booking-Results	♡	What do you think will happen if you click on this icon?		What happened when you clicked on this icon?	
				Why do you think that happened?	
		Why?		Do you think this icon adequately represents its intended function on the interface? Why?	
		Do you recognize the object represented in this icon? If YES, what is that object?		Do you have any suggestions about how this icon could more clearly represent its intended function?	
		How confident are you about this icon interpretation?			
		Which part of Interface was more helpful to understand this icon's function? (icon only, icon and its context - give percentage % in case of both)			
	☝	What do you think will happen if you click on this icon?		What happened when you clicked on this icon?	
				Why do you think that happened?	
		Why?		Do you think this icon adequately represents its intended function on the interface? Why?	
		Do you recognize the object represented in this icon? If YES, what is that object?		Do you have any suggestions about how this icon could more clearly represent its intended function?	
		How confident are you about this icon interpretation?			
		Which part of Interface was more helpful to understand this icon's function? (icon only, icon and its context - give percentage % in case of both)			
	⍐	What do you think will happen if you click on this		What happened when you clicked on this icon?	

		icon?		Why do you think that happened?	
		Why?		Do you think this icon adequately represents its intended function on the interface? Why?	
		Do you recognize the object represented in this icon? If YES, what is that object?		Do you have any suggestions about how this icon could more clearly represent its intended function?	
		How confident are you about this icon interpretation?			
		Which part of Interface was more helpful to understand this icon's function? (icon only, icon and its context - give percentage % in case of both)			

Participant Notes:

-

Appendix-4.2: Large-size screenshots of the Prototype in Study I



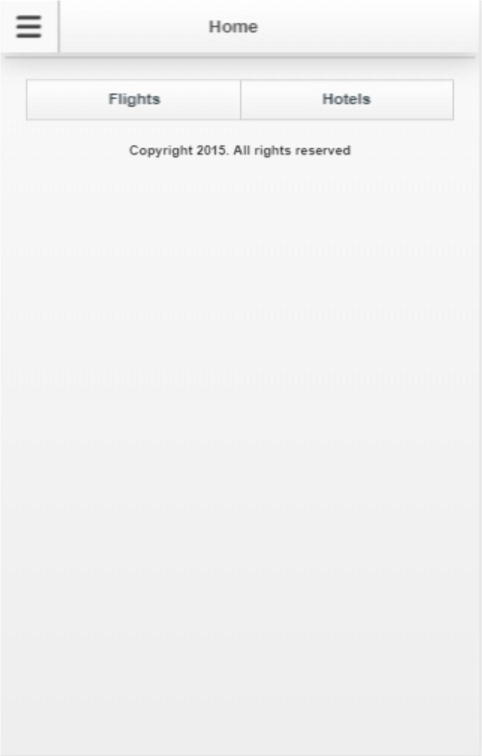
Holiday Booking Website

Sign in

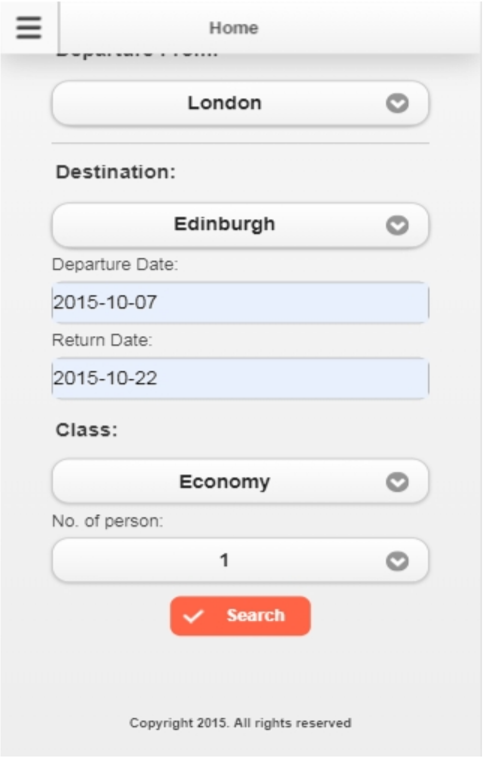
waleed

•••

Copyright 2015. All rights reserved



Copyright 2015. All rights reserved



London

Destination:

Edinburgh

Departure Date:

2015-10-07

Return Date:

2015-10-22

Class:

Economy

No. of person:

1

Copyright 2015. All rights reserved

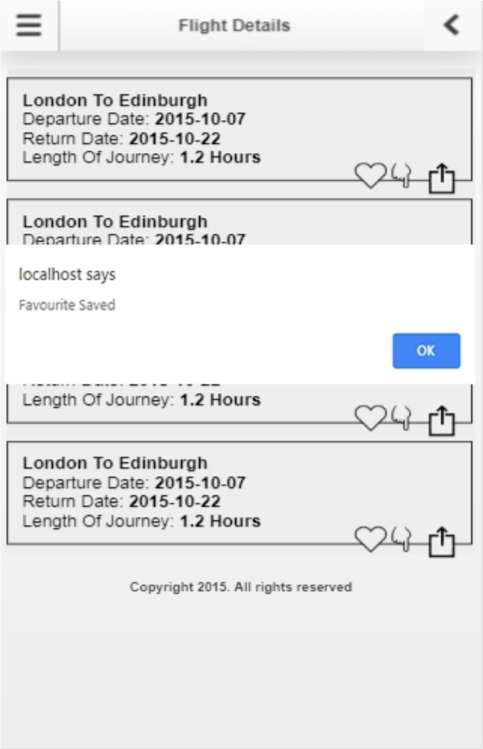
London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.1 Hours**

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**

Copyright 2015. All rights reserved



Flight Details

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**

London To Edinburgh
Departure Date: **2015-10-07**

localhost says

Favourite Saved

OK

Length Of Journey: **1.2 Hours**

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**

Copyright 2015. All rights reserved

Flight Details

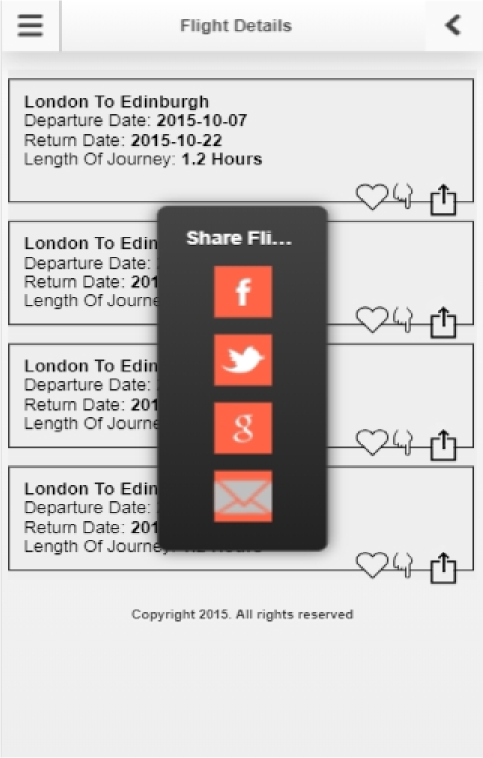
London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**
Price: **125**
Flight Type: **Domestic**

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.1 Hours**

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**

Copyright 2015. All rights reserved



Flight Details

London To Edinburgh
Departure Date: **2015-10-07**
Return Date: **2015-10-22**
Length Of Journey: **1.2 Hours**

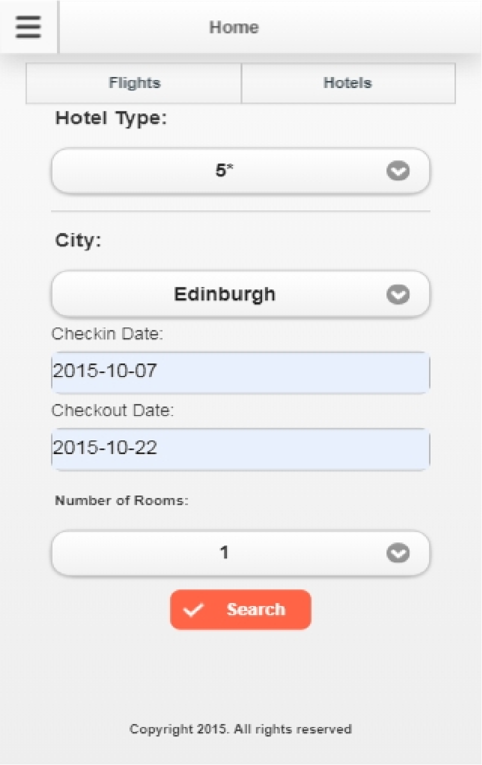
London To Edin
Departure Date:
Return Date: 201
Length Of Journe

London To Edin
Departure Date:
Return Date: 201
Length Of Journe

London To Edin
Departure Date:
Return Date: 201
Length Of Journe

Share Fli...

Copyright 2015. All rights reserved



Hotel Type: 5*

City: Edinburgh

Checkin Date: 2015-10-07

Checkout Date: 2015-10-22

Number of Rooms: 1

Copyright 2015. All rights reserved

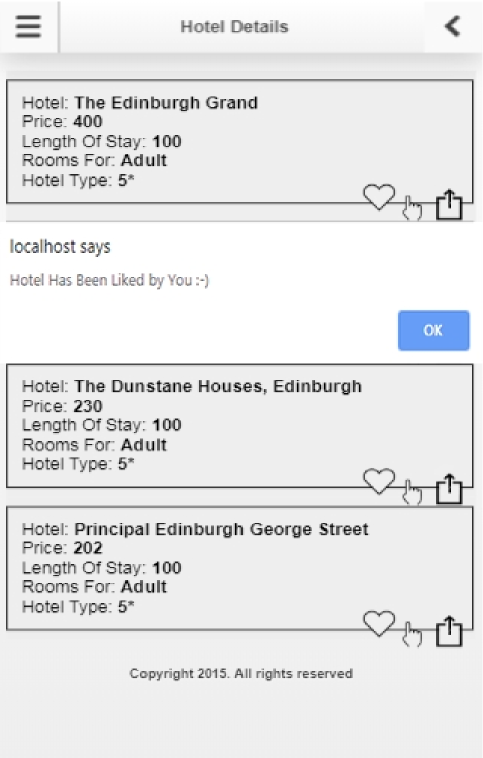
Hotel: **The Edinburgh Grand**
Price: **400**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

Hotel: **The Chester Residence, Edinburgh**
Price: **200**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

Hotel: **The Dunstane Houses, Edinburgh**
Price: **230**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

Hotel: **Principal Edinburgh George Street**
Price: **202**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

Copyright 2015. All rights reserved



Hotel: **The Edinburgh Grand**
Price: **400**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

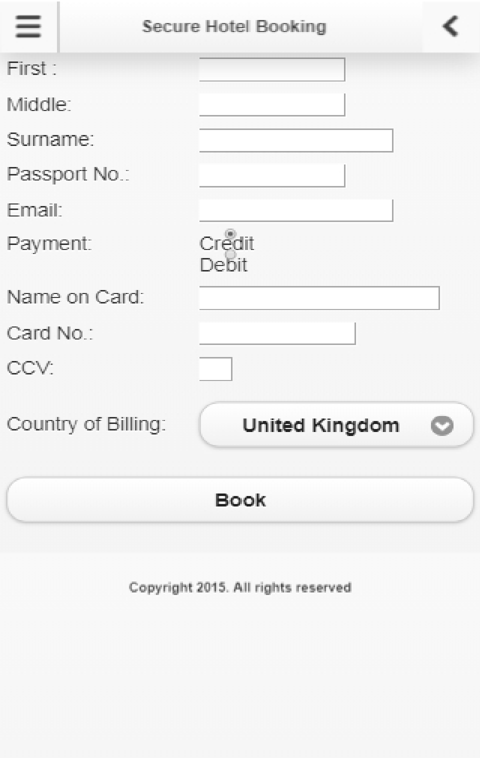
localhost says

Hotel Has Been Liked by You :-)

Hotel: **The Dunstane Houses, Edinburgh**
Price: **230**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

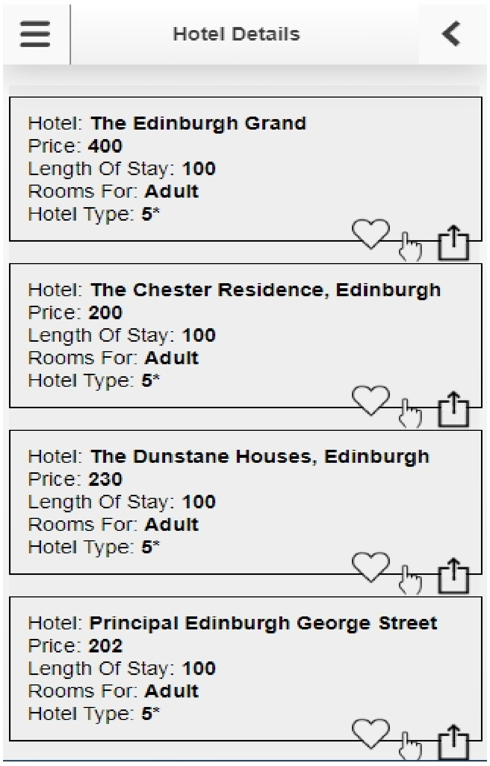
Hotel: **Principal Edinburgh George Street**
Price: **202**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

Copyright 2015. All rights reserved



First :
Middle:
Surname:
Passport No.:
Email:
Payment: ○ Credit ○ Debit
Name on Card:
Card No.:
CCV:
Country of Billing: United Kingdom

Copyright 2015. All rights reserved



Hotel: **The Edinburgh Grand**
Price: **400**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

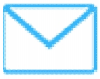
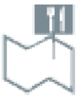
Hotel: **The Chester Residence, Edinburgh**
Price: **200**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

Hotel: **The Dunstane Houses, Edinburgh**
Price: **230**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

Hotel: **Principal Edinburgh George Street**
Price: **202**
Length Of Stay: **100**
Rooms For: **Adult**
Hotel Type: **5***

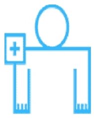
Appendix-5.1:Semantic similarity values using greedy pairing approach

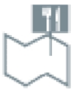
(a) ‘Forward’ presentation order

Family Name	Target Icon	Meaning	Similar Icons	Test description	Semantic Similarity Score
Password		*To show the entry field for the user’s password*		*To show the secret flight offer for Today*	0.57950
				*To get the boarding pass for this flight via payment*	0.52214
				*To get the password to access VIP lounge*	0.58264
E-mail		*To show the entry field for the user’s e-mail address*		*To ask for the discount code via e-mail*	0.69225
				*To get the flight details via e-mail*	0.50363
				*To get the travel checklist via e-mail*	0.52091
User Account		*To add a new user account*		*To upgrade to a premium user account*	0.72285
				*To add a new passenger to the flight*	0.66240
				*To add money to the user’s account*	0.61538
Map		*To show the location of the hotel on the map*		*To show the nearest car parks to the hotel on the map*	0.70588
				*To show the nearest restaurants to the hotel on the map*	0.75000
				*To show the route between the hotel and the city centre on the map*	0.70602
Review		*To show customers’ reviews of the hotel*		*To show customers’ reviews of the car parking services provided by the hotel*	0.76190

				*To show customers' reviews of room service in the hotel*	0.84211
				*To show customers' reviews of the staff in the hotel*	0.88889
Like		*To show the number of customers who liked the hotel*		*To show the number of customers who liked the car park services provided by the hotel*	0.78261
				*To show the number of customers who liked the food provided in the hotel*	0.85714
				*To show the number of customers who liked the cleanliness of the hotel*	0.94737

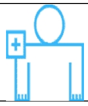
(b) 'Backward' presentation order

Family Name	Target Icon	Meaning	Similar Icons	Test description	Semantic Similarity Score
Password		*To show the entry field for the user's password*		*To show the secret flight offer for Today*	*0.57950*
				*to get the boarding pass for this flight via payment*	*0.52214*
				*To get the password to access VIP lounge*	*0.58264*
E-mail		*To show the entry field for the user's e-mail address*		*to ask for the discount code via e-mail*	0.69225
				*to get the flight details via e-mail*	0.50363
				*to get the travel checklist via e-mail*	0.52091
User Account		*To add a new user account*		*To upgrade to a premium user account*	0.72285
				*to add a new passenger to the flight*	0.66240

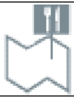
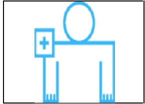
				*To add money to the user's account*	0.61538
Map		*To show the location of the hotel on the map*		*To show the nearest car parking to the hotel on the map*	0.70588
				*To show the nearest restaurants to the hotel on the map*	0.75000
				*To show the route between the hotel and the city centre on the map*	0.70602
Review		*To show customers' reviews of the hotel*		*To show customers' reviews of the car parking services provided by the hotel*	0.76190
				*To show customers' reviews of room service in the hotel*	0.84211
				*To show customers' reviews of the staff in the hotel*	0.88889
Like		*To show the number of customers who liked the hotel*		*To show the number of customers who liked the car park services provided by the hotel*	0.78261
				*To show the number of customers who liked the food provided in the hotel*	0.85714
				*To show the number of customers who liked the cleanliness of the hotel*	0.94737

Appendix-5.2: Semantic similarity of randomly selected icons in different families

(a) 'Forward' presentation order

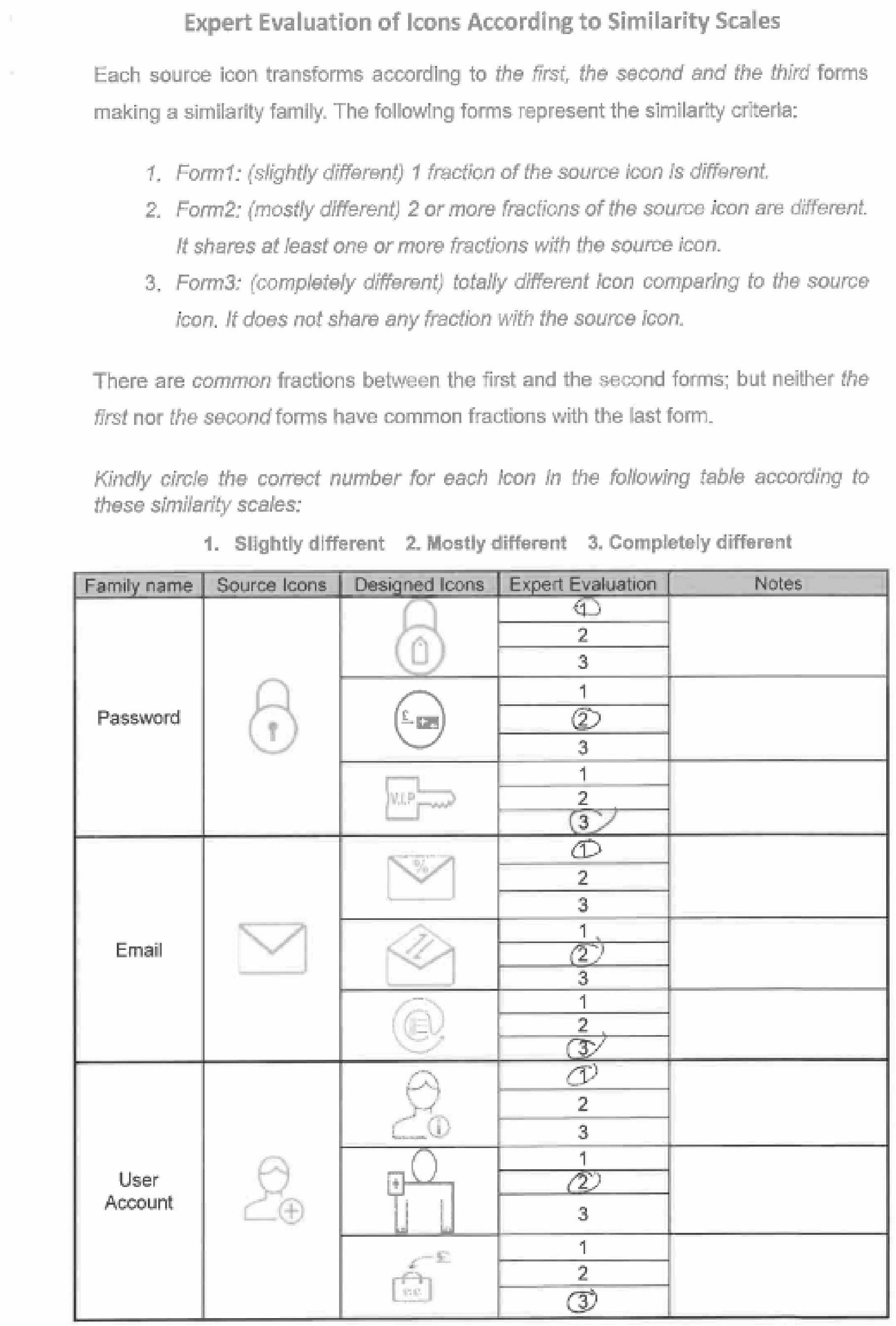
Icon	Meaning	Family Name	Icon	Meaning	Family Name	Semantic Similarity Value
	*To show customers' reviews of the staff in this hotel*	*Review*		*To show the nearest restaurants to the hotel on the map*	*Map*	0.44444
	*To show the number of customers who liked the food provided in this hotel*	*Like*		*To show the route between the hotel and the city centre on the map*	*Map*	0.39969
	*to send the flight details to the user via e-mail*	*Email*		*To show customers' reviews of room service in this hotel*	*Review*	*0.24561*
	To show the location of the hotel on the map	*Map*		*To show customers' reviews of room service in this hotel*	*Review*	*0.42105*
	*To show the entry field for the user's password*	*Password*		*To upgrade to a premium user account*	*User account*	*0.37347*
	*To show the entry field for the user's e-mail address*	*Email*		*to get the boarding pass for this flight via payment*	*Password*	*0.46056*
	*To get the password to access VIP lounge*	*Password*		*To show the entry field for the user's e-mail address*	*Email*	*0.43072*
	*to send the travel checklist to the user via e-mail*	*Email*		*To show the secret flight offer for today*	*Password*	*0.47660*
	*To add a new passenger to the flight*	*User Account*		*To show the entry field for the user's password*	*Password*	*0.32707*
	*To show the number of customers who liked the car park services provided by this hotel*	*Like*		*To upgrade to a premium user account*	*User Account*	*0.15964*
	*To show the location of the hotel on the map*	*Map*		*to add a new passenger to the flight*	*User Account*	*0.26667*
	*To show the nearest car parking to the hotel on the map*	*Map*		*To show customers' reviews of room service in this hotel*	*Review*	0.43893
	*To show customers' reviews about the car parking services provided by the hotel*	*Review*		*To show the route between the hotel and the city centre on the map*	*Map*	0.38742

(b) ‘Backward’ presentation order

Icon	Meaning	Family Name	Icon	Meaning	Family Name	Semantic Similarity Value
	*To show customers’ reviews of the staff in this hotel*	Review		*To show the nearest restaurants to the hotel on the map*	Map	0.44444
	*To show the number of customers who liked the food provided in this hotel*	Like		*To show the route between the hotel and the city centre on the map*	Map	0.39969
	*to send the flight details to the user via e-mail*	Email		*To show customers’ reviews of room service in this hotel*	Review	0.24561
	*To show the location of the hotel on the map*	Map		*To show customers’ reviews of room service in this hotel*	Review	0.42105
	*To show the entry field for the user’s password*	Password		*To upgrade to a premium user account*	User account	0.37347
	*To show the entry field for the user’s e-mail address*	Email		*to get the boarding pass for this flight via payment*	Password	0.46056
	*To get the password to access VIP lounge*	Password		*To show the entry field for the user’s e-mail address*	Email	0.43072
	*to send the travel checklist to the user via e-mail*	Email		*To show the secret flight offer for today*	Password	0.47660
	*To add a new passenger to the flight*	User Account		*To show the entry field for the user’s password*	Password	0.32707
	*To show the number of customers who liked the car park services provided by this hotel*	Like		*To upgrade to a premium user account*	User Account	0.15964
	*To show the location of the hotel on the map*	Map		*to add a new passenger to the flight*	User Account	0.26667
	*To show the nearest car parking to the hotel on the map*	Map		*To show customers’ reviews of room service in this hotel*	Review	0.43893
	*To show customers’ reviews about the car parking services provided by the hotel*	Review		*To show the route between the hotel and the city centre on the map*	Map	0.38742

Appendix-5.3: Expert test of visual similarity ('forward' presentation order)

(a) Responses of the first expert:



Expert Evaluation of Icons According to Similarity Scales

Each source icon transforms according to *the first, the second and the third* forms making a similarity family. The following forms represent the similarity criteria:

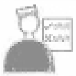
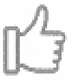
1. *Form1: (slightly different) 1 fraction of the source icon is different.*
2. *Form2: (mostly different) 2 or more fractions of the source icon are different. It shares at least one or more fractions with the source icon.*
3. *Form3: (completely different) totally different icon comparing to the source icon. It does not share any fraction with the source icon.*

There are *common* fractions between the first and the second forms; but neither *the first* nor *the second* forms have common fractions with the last form.

Kindly circle the correct number for each icon in the following table according to these similarity scales:

1. Slightly different 2. Mostly different 3. Completely different

Family name	Source Icons	Designed Icons	Expert Evaluation	Notes
Password			[✓ 1]	
			2	
			3	
			1	
			[✓ 2]	
			3	
			1	
			2	
			[✓ 3]	
Email			[✓ 1]	
			2	
			3	
			1	
			[✓ 2]	
			3	
			1	
			2	
			[✓ 3]	
User Account			[✓ 1]	
			2	
			3	
			1	
			[✓ 2]	
			3	
			1	
			2	
			[✓ 3]	

Map			[✓ 1] 2 3	
			1 [✓ 2] 3	
			1 2 [✓ 3]	
Reviews			[✓ 1] 2 3	
			1 [✓ 2] 3	
			1 2 [✓ 3]	
Like			[✓ 1] 2 3	
			1 [✓ 2] 3	
			1 2 [✓ 3]	

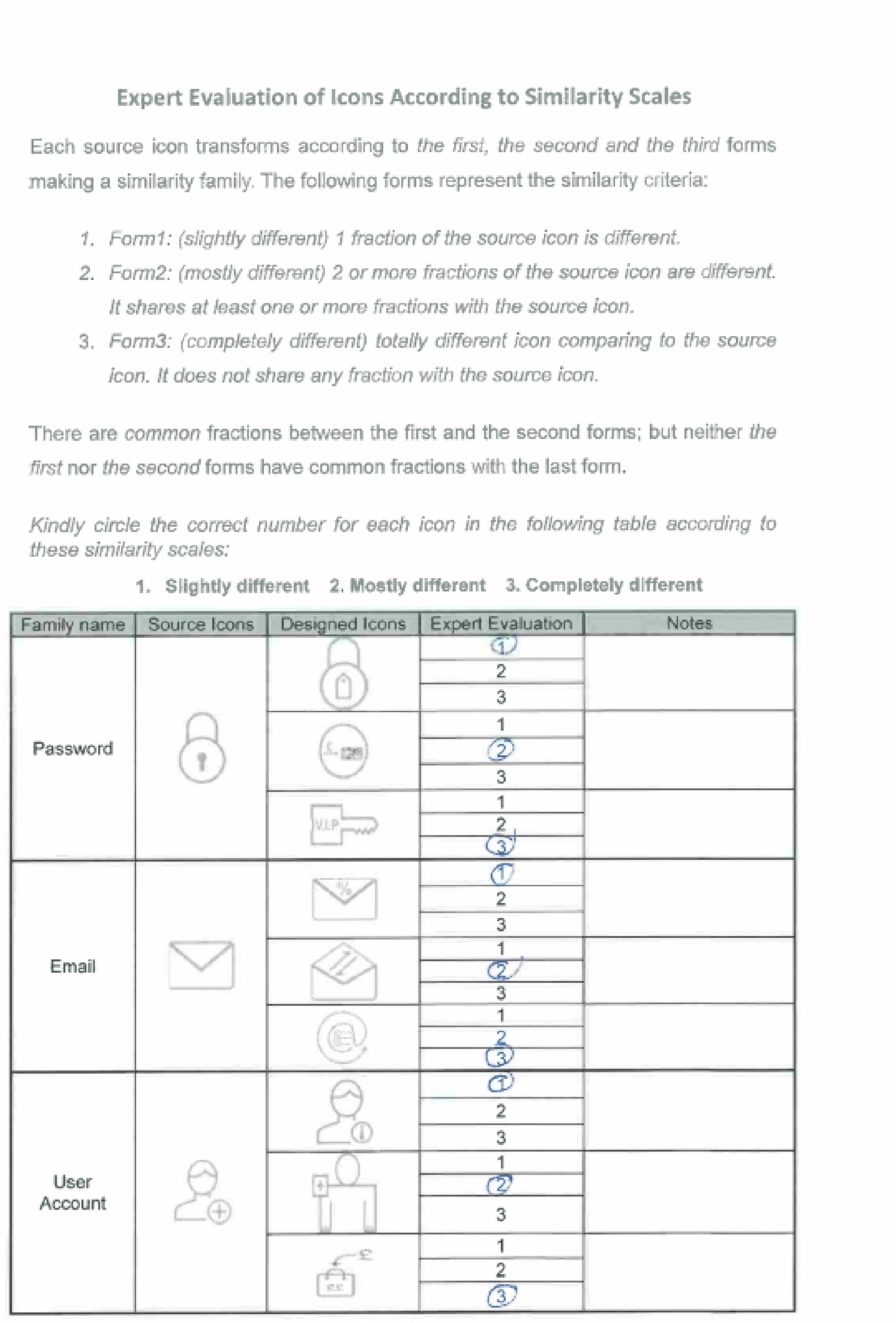
Full Name: [redacted]

Academic or Professional Qualifications: [redacted]

Current Position: [redacted]

Contact(Mobile/Email): [redacted]

(b) Responses of the second expert:



Expert Evaluation of Icons According to Similarity Scales

Each source icon transforms according to *the first, the second and the third* forms making a similarity family. The following forms represent the similarity criteria:

1. *Form1: (slightly different) 1 fraction of the source icon is different.*
2. *Form2: (mostly different) 2 or more fractions of the source icon are different. It shares at least one or more fractions with the source icon.*
3. *Form3: (completely different) totally different icon comparing to the source icon. It does not share any fraction with the source icon.*

There are *common* fractions between the first and the second forms; but neither *the first* nor *the second* forms have common fractions with the last form.

Kindly circle the correct number for each icon in the following table according to these similarity scales:

1. Slightly different 2. Mostly different 3. Completely different

Family name	Source Icons	Designed Icons	Expert Evaluation	Notes
Password			[✓ 1] 2 3	
			1 [✓ 2] 3	
		V.I.P	1 2 [✓ 3]	
Email			[✓ 1] 2 3	
			1 [✓ 2] 3	
			1 2 [✓ 3]	
User Account			[✓ 1] 2 3	
			1 [✓ 2] 3	
			1 2 [✓ 3]	

Map	🗺	P 🗺	[✓ 1] 2 3	
		🗺	1 [✓ 2] 3	
		🗺	1 2 [✓ 3]	
Reviews	★★★🔍	★★★ P 🔍	[✓ 1] 2 3	
		🔍	1 [✓ 2] 3	
			1 2 [✓ 3]	
Like	👍	P 👍	[✓ 1] 2 3	
		👍	1 [✓ 2] 3	
		♥	1 2 [✓ 3]	

Full Name: [redacted]

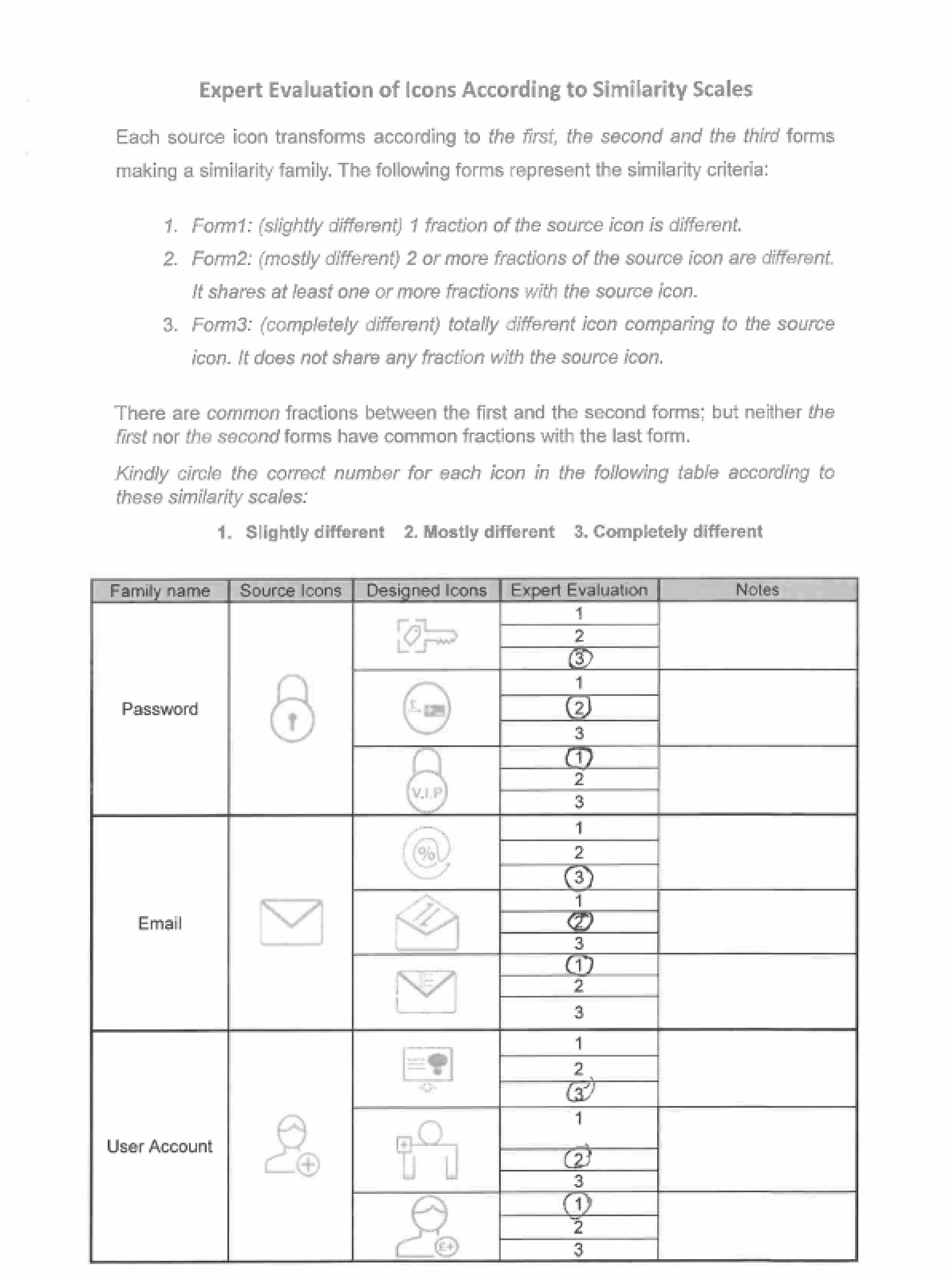
Academic or Professional Qualifications:

Current Position: [redacted]

Cont [redacted]

Appendix-5.4: Expert test of visual similarity (‘backward’ presentation order)

(a) Responses of the first expert:



Expert Evaluation of Icons According to Similarity Scales

Each source icon transforms according to *the first, the second and the third* forms making a similarity family. The following forms represent the similarity criteria:

1. *Form1: (slightly different) 1 fraction of the source icon is different.*
2. *Form2: (mostly different) 2 or more fractions of the source icon are different. It shares at least one or more fractions with the source icon.*
3. *Form3: (completely different) totally different icon comparing to the source icon. It does not share any fraction with the source icon.*

There are *common* fractions between the first and the second forms; but neither *the first* nor *the second* forms have common fractions with the last form.

Kindly circle the correct number for each icon in the following table according to these similarity scales:

1. Slightly different 2. Mostly different 3. Completely different

Family name	Source Icons	Designed Icons	Expert Evaluation	Notes
Password			1 2 [✓ 3]	
			1 [✓ 2] 3	
		V.I.P	[✓ 1] 2 3	
Email		%	1 2 [✓ 3]	
			1 [✓ 2] 3	
			[✓ 1] 2 3	
User Account			1 2 [✓ 3]	
			1 [✓ 2] 3	
			[✓ 1] 2 3	

Map			1 2 [✓ 3]	
			1 [✓ 2] 3	
			[✓ 1] 2 3	
Reviews			1 2 [✓ 3]	
			1 [✓ 2] 3	
			[✓ 1] 2 3	
Like			1 2 [✓ 3]	
			1 [✓ 2] 3	
			[✓ 1] 2 3	

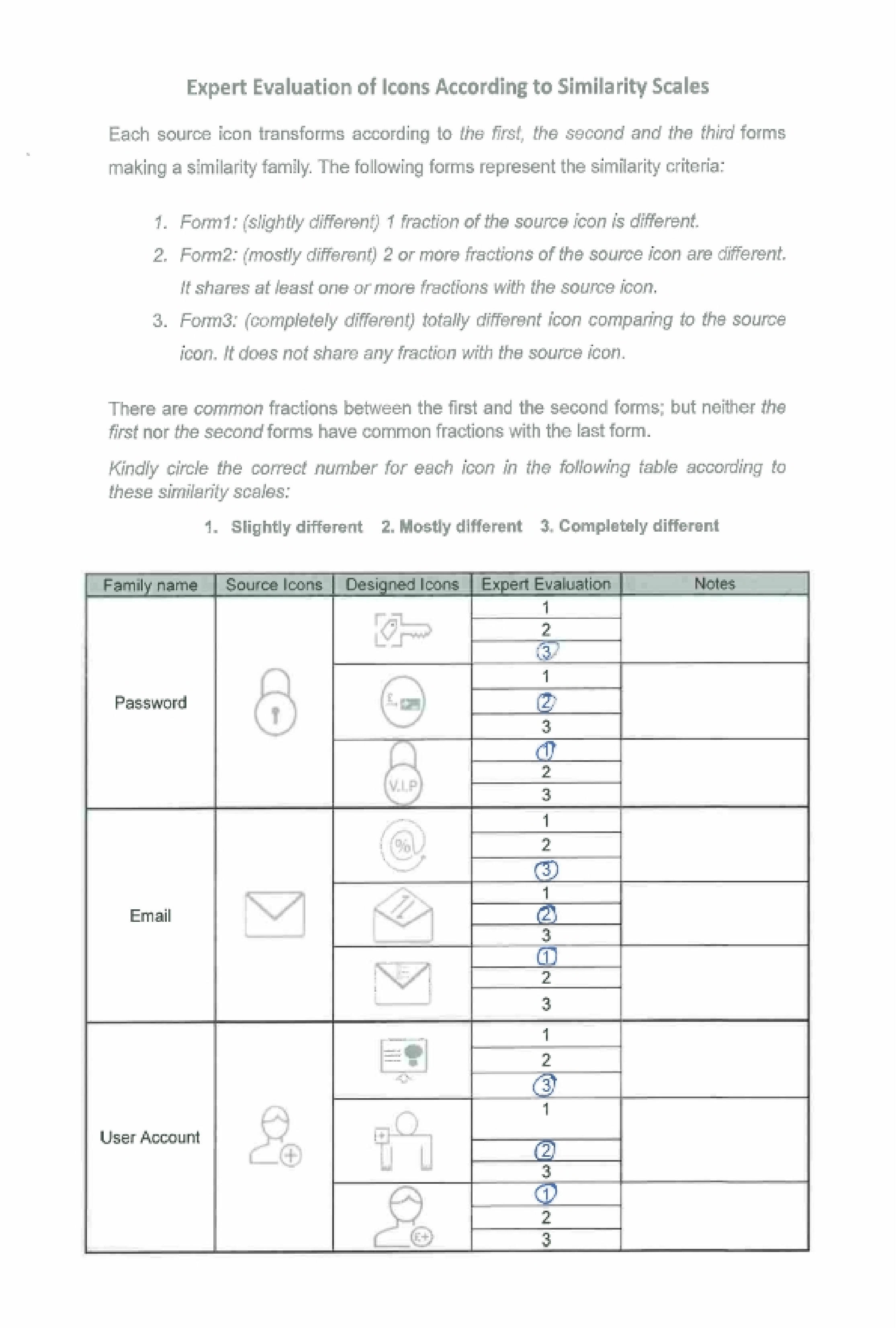
Full Name: [redacted]

Academic or Professional Qualifications: [redacted]

Current Position: [redacted]

Contact(Mobile/Email): [redacted]

(b) Responses of the second expert:



Expert Evaluation of Icons According to Similarity Scales

Each source icon transforms according to *the first, the second and the third* forms making a similarity family. The following forms represent the similarity criteria:

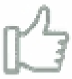
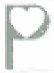
1. *Form1: (slightly different) 1 fraction of the source icon is different.*
2. *Form2: (mostly different) 2 or more fractions of the source icon are different. It shares at least one or more fractions with the source icon.*
3. *Form3: (completely different) totally different icon comparing to the source icon. It does not share any fraction with the source icon.*

There are *common* fractions between the first and the second forms; but neither *the first* nor *the second* forms have common fractions with the last form.

Kindly circle the correct number for each icon in the following table according to these similarity scales:

1. Slightly different 2. Mostly different 3. Completely different

Family name	Source Icons	Designed Icons	Expert Evaluation	Notes
Password			1	
			2	
			[✓ 3]	
			1	
			[✓ 2]	
			3	
		V.I.P	[✓ 1]	
			2	
			3	
Email		%	1	
			2	
			[✓ 3]	
			1	
			[✓ 2]	
			3	
			[✓ 1]	
			2	
			3	
User Account			1	
			2	
			[✓ 3]	
			1	
			[✓ 2]	
			3	
		E+	[✓ 1]	
			2	
			3	

Map			1 2 [✓ 3]	
			1 [✓ 2] 3	
			[✓ 1] 2 3	
Reviews			1 2 [✓ 3]	
			1 [✓ 2] 3	
			[✓ 1] 2 3	
Like			1 2 [✓ 3]	
			1 [✓ 2] 3	
			[✓ 1] 2 3	

Full Name: [redacted]

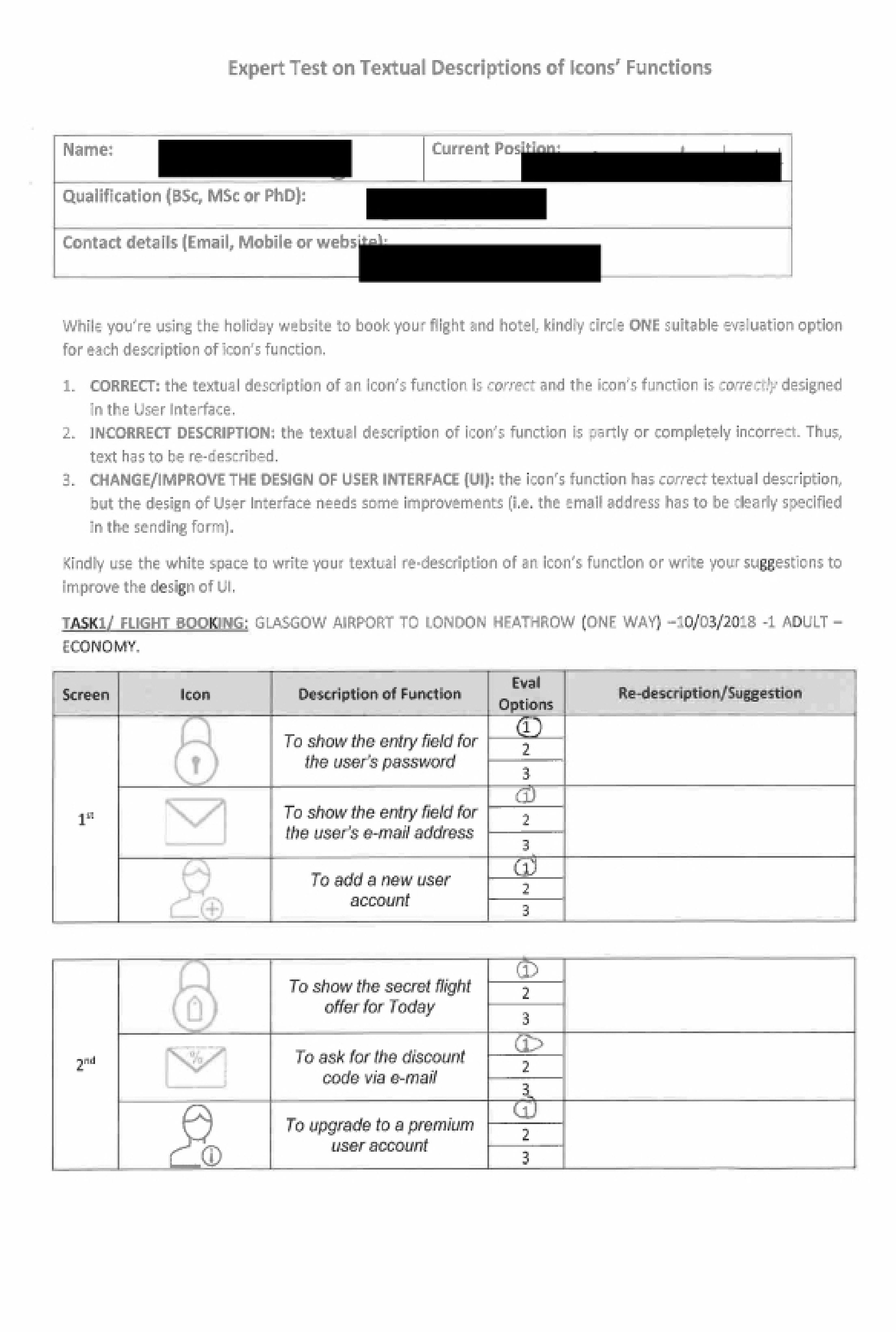
Academic or Professional Qualifications:

Current Position: [redacted]

Co[redacted]

Appendix-5.5: Expert response to the expert test of the textual descriptions of the icons

(a) Responses of the first expert:



Expert Test on Textual Descriptions of Icons' Functions

Name: [redacted]

Current Position: [redacted]

Qualification (BSc, MSc or PhD): [redacted]

Contact details (Email, Mobile or website): [redacted]

While you're using the holiday website to book your flight and hotel, kindly circle **ONE** suitable evaluation option for each description of icon's function.

1. **CORRECT:** the textual description of an icon's function is *correct* and the icon's function is *correctly* designed in the User Interface.
2. **INCORRECT DESCRIPTION:** the textual description of icon's function is partly or completely incorrect. Thus, text has to be re-described.
3. **CHANGE/IMPROVE THE DESIGN OF USER INTERFACE (UI):** the icon's function has *correct* textual description, but the design of User Interface needs some improvements (i.e. the email address has to be clearly specified in the sending form).

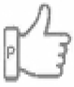
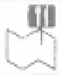
Kindly use the white space to write your textual re-description of an icon's function or write your suggestions to improve the design of UI.

<u>TASK1/ FLIGHT BOOKING:</u> GLASGOW AIRPORT TO LONDON HEATHROW (ONE WAY) –10/03/2018 -1 ADULT – ECONOMY.

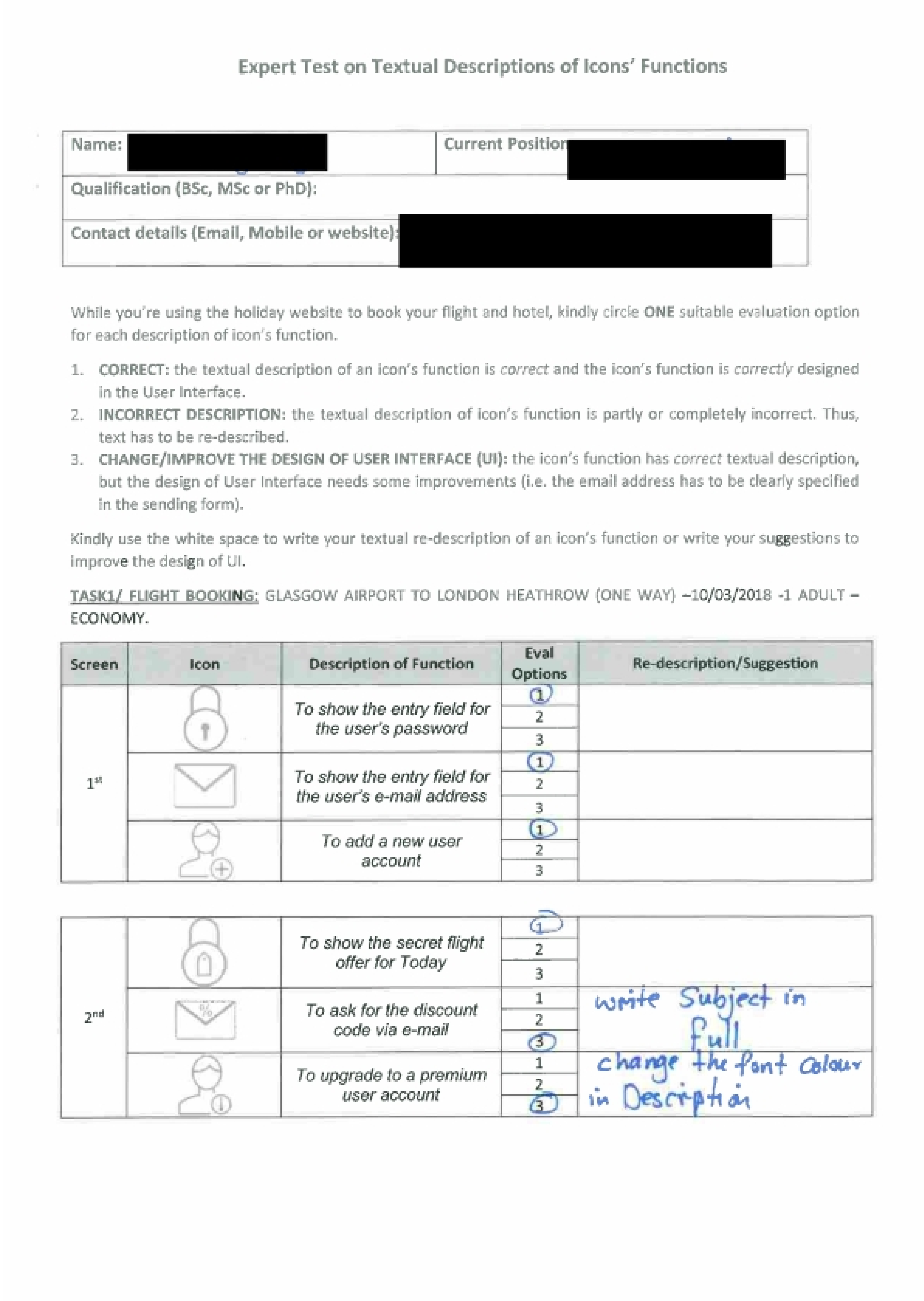
Screen	Icon	Description of Function	Eval Options	Re-description/Suggestion
1st	🔒	*To show the entry field for the user's password*	[✓ 1] 2 3	
	✉	*To show the entry field for the user's e-mail address*	[✓ 1] 2 3	
	👤	*To add a new user account*	[✓ 1] 2 3	
2nd	🔒	*To show the secret flight offer for Today*	[✓ 1] 2 3	
	✉	*To ask for the discount code via e-mail*	[✓ 1] 2 3	
	👤	*To upgrade to a premium user account*	[✓ 1] 2 3	

Screen	Icon	Description	Eval Options	Re-description/Suggestion
3rd		*To get the boarding pass for this flight via payment*	[✓ 1] 2 3	
		*To get the flight details via e-mail*	[✓ 1] 2 3	
		*To add a new passenger to the flight*	[✓ 1] 2 3	
4th		*To get the password to access V.I.P lounge*	[✓ 1] 2 3	
		*To get the travel checklist via e-mail*	[✓ 1] 2 3	why not the email folder?
		*To add money to the user's account*	[✓ 1] 2 3	user's ≡ your

TASK2/ HOTEL BOOKING: LONDON - 10/03/2018 to 14-03-2018- 1 ADULT- 1 ROOM.

Screen	Icon	Description	Eval Options	Re-description/Suggestion
1st		*To show the location of the hotel on the map*	[✓ 1] 2 3	
		*To show customers' reviews of the hotel*	[✓ 1] 2 3	
		*To show the number of customers who liked the hotel*	[✓ 1] 2 3	
2nd		*To show the nearest car parks to the hotel on the map*	[✓ 1] 2 3	
		*To show customers' reviews of the car parking services provided by the hotel*	[✓ 1] 2 3	
		*To show the number of customers who liked the car park services provided by the hotel*	[✓ 1] 2 3	
3rd		*To show the nearest restaurants to the hotel on the map*	[✓ 1] 2 3	
		*To show customers' reviews of room service in the hotel*	[✓ 1] 2 3	
		*To show the number of customers who liked the food provided in the hotel*	[✓ 1] 2 3	
4th		*To show the route between the hotel and the city center on the map*	[✓ 1] 2 3	
		*To show customers' reviews of the staff in the hotel*	[✓ 1] 2 3	why change the "Search" magnify
		*To show the number of customers who liked the cleanliness of the hotel*	[✓ 1] 2 3	why change the thumb-up?

(b) Responses of the second expert:



Expert Test on Textual Descriptions of Icons' Functions

Name: [redacted]	**Current Position:** [redacted]
Qualification (BSc, MSc or PhD):	
Contact details (Email, Mobile or website): [redacted]	

While you're using the holiday website to book your flight and hotel, kindly circle **ONE** suitable evaluation option for each description of icon's function.

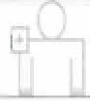
1. **CORRECT:** the textual description of an icon's function is *correct* and the icon's function is *correctly* designed in the User Interface.
2. **INCORRECT DESCRIPTION:** the textual description of icon's function is partly or completely incorrect. Thus, text has to be re-described.
3. **CHANGE/IMPROVE THE DESIGN OF USER INTERFACE (UI):** the icon's function has *correct* textual description, but the design of User Interface needs some improvements (i.e. the email address has to be clearly specified in the sending form).

Kindly use the white space to write your textual re-description of an icon's function or write your suggestions to improve the design of UI.

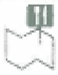
<u>TASK1/ FLIGHT BOOKING:</u> GLASGOW AIRPORT TO LONDON HEATHROW (ONE WAY) –10/03/2018 -1 ADULT – ECONOMY.

Screen	Icon	Description of Function	Eval Options	Re-description/Suggestion
1st	🔒	*To show the entry field for the user's password*	[✓ 1] 2 3	
	✉	*To show the entry field for the user's e-mail address*	[✓ 1] 2 3	
	👤⊕	*To add a new user account*	[✓ 1] 2 3	

2nd	🔒	*To show the secret flight offer for Today*	[✓ 1] 2 3	
	✉%	*To ask for the discount code via e-mail*	1 2 [✓ 3]	write Subject in full
	👤ⓘ	*To upgrade to a premium user account*	1 2 [✓ 3]	change the font colour in Description

Screen	Icon	Description	Eval Options	Re-description/Suggestion
3rd		*To get the boarding pass for this flight via payment*	[✓ 1]	
			2	
			3	
		*To get the flight details via e-mail*	[✓ 1]	
			2	
			3	
		*To add a new passenger to the flight*	[✓ 1]	
			2	
			3	
4th		*To get the password to access V.I.P lounge*	1	Change the font colour
			2	
			[✓ 3]	
		*To get the travel checklist via e-mail*	[✓ 1]	
			2	
			3	
		*To add money to the user's account*	[✓ 1]	
			2	
			3	

TASK2/ HOTEL BOOKING: LONDON - 10/03/2018 to 14-03-2018- 1 ADULT- 1 ROOM.

Screen	Icon	Description	Eval Options	Re-description/Suggestion
1st		*To show the location of the hotel on the map*	1 2 [✓ 3]	The UI could be improved the font size in description is very large.
		*To show customers' reviews of the hotel*	[✓ 1] 2 3	
		*To show the number of customers who liked the hotel*	[✓ 1] 2 3	
2nd		*To show the nearest car parks to the hotel on the map*	[✓ 1] 2 3	
		*To show customers' reviews of the car parking services provided by the hotel*	[✓ 1] 2 3	
		*To show the number of customers who liked the car park services provided by the hotel*	[✓ 1] 2 3	
3rd		*To show the nearest restaurants to the hotel on the map*	[✓ 1] 2 3	
		*To show customers' reviews of room service in the hotel*	[✓ 1] 2 3	
		*To show the number of customers who liked the food provided in the hotel*	[✓ 1] 2 3	
4th		*To show the route between the hotel and the city center on the map*	[✓ 1] 2 3	
		*To show customers' reviews of the staff in the hotel*	[✓ 1] 2 3	
		*To show the number of customers who liked the cleanliness of the hotel*	[✓ 1] 2 3	

Appendix-5.6: Guidance Sheet (The Hotel-First Group)

Hello, and thanks for accepting my invitation to take part in this study. If you can kindly let me know your **name**.

Before we start, I will introduce you with an overview of the study. There are five main phases:

1. **Calibration:** You need to **calibrate** your eyes with the eye-tracker by following the instructions. We might need to do this several times until we get your eyes well-calibrated with the display monitor.
2. **Coaching for Task 1:** You will start a coaching session for task 1.
3. **Task 1:** You will be requested to make a hotel booking for your Easter holiday through the holiday booking website.
4. **Coaching for Task 2**: You will start a coaching session for task 2.
5. **Task 2:** You will be requested to make a flight booking for your Easter holiday through the holiday booking website.

Three-step protocol:

For all the coaching and the tasks sessions, you need to follow the *three-step* protocol on each screen:

4. **Viewing in silence:** look at all the **ICONS** IN SILENCE and try to think of what would happen if you clicked on them? When you finish just say 'I AM DONE'. On each main screen, there are three icons with other buttons to proceed with the task (i.e. search, login, and pay buttons).
5. **Before clicking**: For each icon, you need to tell me accurately:
   - What would happen if you click on the icon?
   - Why do you think that would happen?
6. **After clicking on** each icon, tell me! What happened when you clicked on that icon? When you click on an icon, this might take you to another page or open a dialogue box, you don't need to proceed further; just tell me what happened and then close the dialogue box or go back to the previous page.

Information you need during the test

(Please read this information before the real test starts and remember that you will need them during the test)

There are two tasks you need to do in this study:

1. **Hotel Booking Task:**
   For searching for a hotel, please use the following:
   London, from 01-04-2018 until 10-04-2018, 1 ADULT, 1 ROOM
   Kindly complete the task until you successfully pay to secure your booking.

2. **Flight Booking Task:**
   Please note that flight booking task will ask for **login** information

   (Email ID: user@gmail.com, Password: 12345).

   For searching for a flight, please use the following search information:

   One Way, Glasgow Airport to London Heathrow, 01-04-2018, ECONOMY, 1 ADULT

   Kindly complete the task until you successfully pay to secure your booking.

Appendix-5.7: Guidance Sheet (The Flight-First Group)

Hello, and thanks for accepting my invitation to take part in my study. If you can kindly let me know your **name**.

Before we start, I will introduce you with an overview of the study. There are five main phases:

6. **Calibration:** You need to **calibrate** your eyes with the eye-tracker by following the instructions. We might need to do this several times until we get your eyes well-calibrated with the display monitor.
7. **Coaching for Task 1:** You will start a coaching session for task 1.
8. **Task 1:** You will be requested to make a flight booking for your Easter holiday through the holiday booking website.
9. **Coaching for Task 2**: You will start a coaching session for task 2.
10. **Task 2:** You will be requested to make a hotel booking for your Easter holiday through the holiday booking website.

Three-step protocol:

For all the tasks, you need to follow the *three-step* protocol on each screen:

7. **Viewing in silence:** look at all the **ICONS** IN SILENCE and try to think of what would happen if you clicked on them? When you finish just say 'I AM DONE'. In each main screen, there are 3 icons with other buttons to proceed with the task (i.e. search, login and pay buttons).
8. **Before clicking**: For each icon, you need to tell me accurately:
   - What would happen if you click on the icon?
   - Why do you think that would happen?
9. **After clicking on** each icon, tell me! What happened when you clicked on that icon? When you click on an icon, this might take you to another page or open a dialogue box, you don't need to proceed further; just tell me what happened and then close the dialogue box or go backward to the previous page.

Information you need during the test

(Please read this information before the real test starts and remember that you will need them during the test)

There are two tasks you need to do in this study:

3. **Flight Booking Task:**

   Please note that flight booking task will ask for **login** information

   (Email ID: user@gmail.com, Password: 12345).

   For searching for a flight, please use the following search information:

   One Way, Glasgow Airport to London Heathrow, 01-04-2018, ECONOMY, 1 ADULT

   Kindly complete the task until you successfully pay to secure your booking.

2. **Hotel Booking Task:**

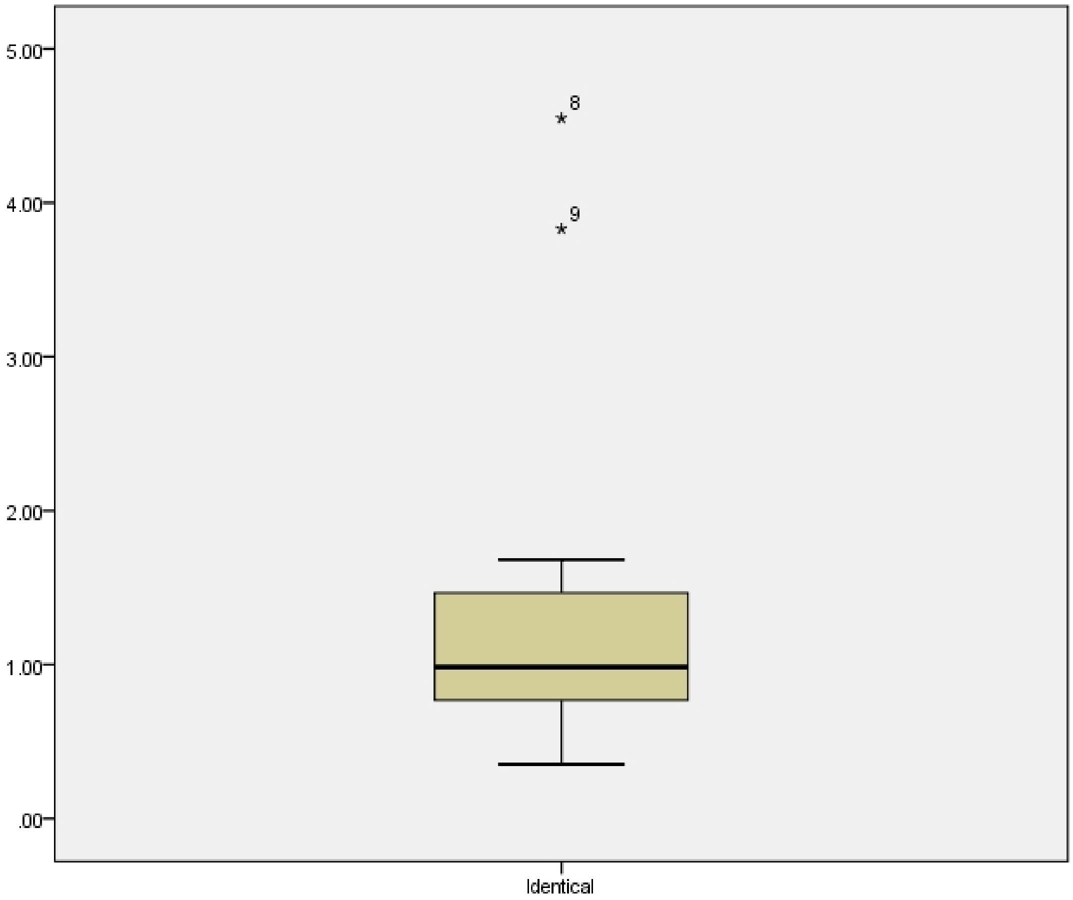
   For searching for a hotel, please use the following:
   London, from 01-04-2018 until 10-04-2018, 1 ADULT, 1 ROOM
   Kindly complete the task until you successfully pay to secure your booking.

Appendix-5.8: Eye-tracking instructions

Instructions for participants:

1. Make sure you sit comfortably.
2. During the calibration phase, make sure that you look at the crossing points. Look for **5 secs** at the crossing point; then click the mouse; then look for **2 secs** before you move your eyes to the next crossing point.
3. Make sure that you don't move your head while looking in silence. Your head must be stable and straight.
4. Make sure you don't use the keyboard or the mouse while looking in silence.
5. Make sure you say your interpretation as accurately as possible.
6. There are three icons require your attention apart from the tabs and the buttons (login, search or moving forward and backward buttons).
7. Please provide precise reasons for the **why** question. For example, because you have experienced this icon on other websites or this website or maybe because of the shape of the icon.
8. You don't need to follow the three-step protocol on the screens that don't have icons.
9. Flight booking requires authentications: Email ID and Password to Log-In.
10. Make sure that you look specifically at the icons during the free viewing phase.

Appendix-5.9: RQ2.1 Outliers Boxplot



Appendix-5.9: Boxplot for RQ2.1 showing the speed of recognition of the participants for the 'backward' presentation order, identical icons. Participant 8 is a significant outlier and participant 9 is a nearly significant outlier, as indicated by z-scores.

Appendix-5.10: RQ2.1 Shapiro-Wilk Normality Test

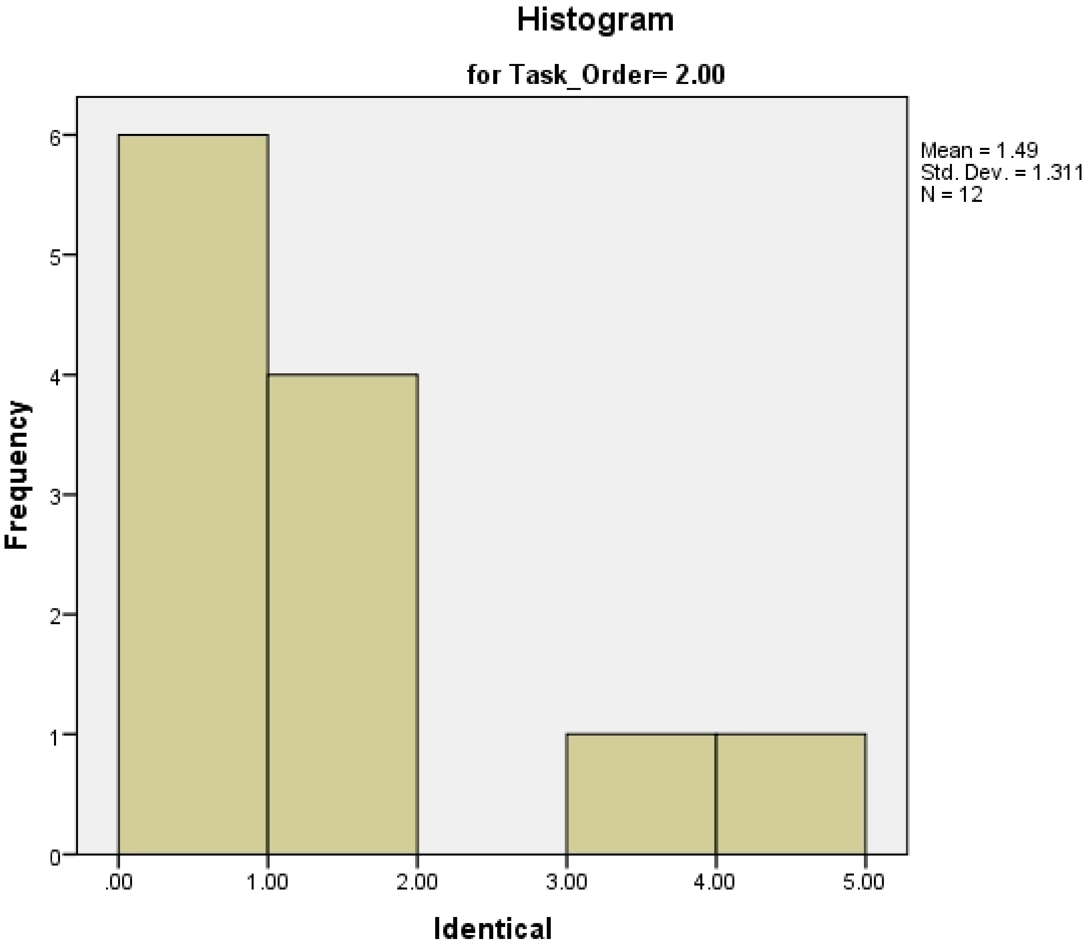
		Tests of Normality		
Icon type	**Presentation order**	**Shapiro-Wilk**		
		Statistic	**df**	**Sig.**
Identical	Forward	.899	12	.154
	Backward	.715	12	.001
Slightly different	Forward	.875	12	.076
	Backward	.918	12	.274
Mostly different	Forward	.955	12	.713
	Backward	.901	12	.164
Dissimilar	Forward	.981	12	.987
	Backward	.904	12	.178

*. This is a lower bound of the true significance.

a. Lilliefors significance correction

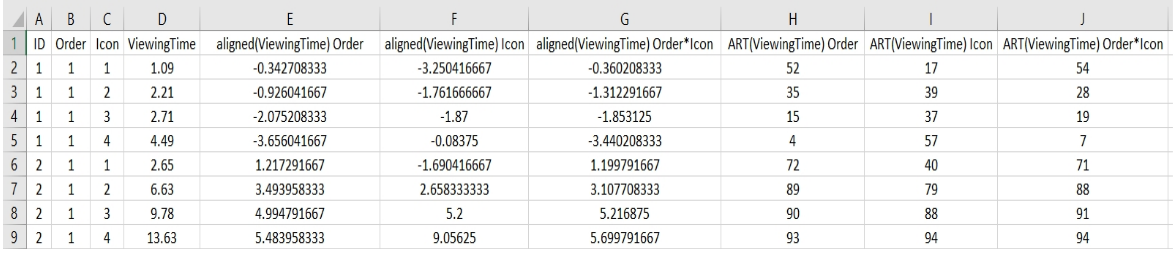
Appendix-5.10: The Shapiro-Wilk test of normality for RQ2.1. The ‘backward’ presentation, identical icon condition has a p-value below 0.05 (in red), which indicates that the data are not normally distributed for this condition.

Appendix-5.11: RQ2.1 Histogram



Appendix-5.11: Histogram for RQ2.1 showing the distribution of the participants' speed of recognition times in the 'backward' presentation order, identical icon condition.

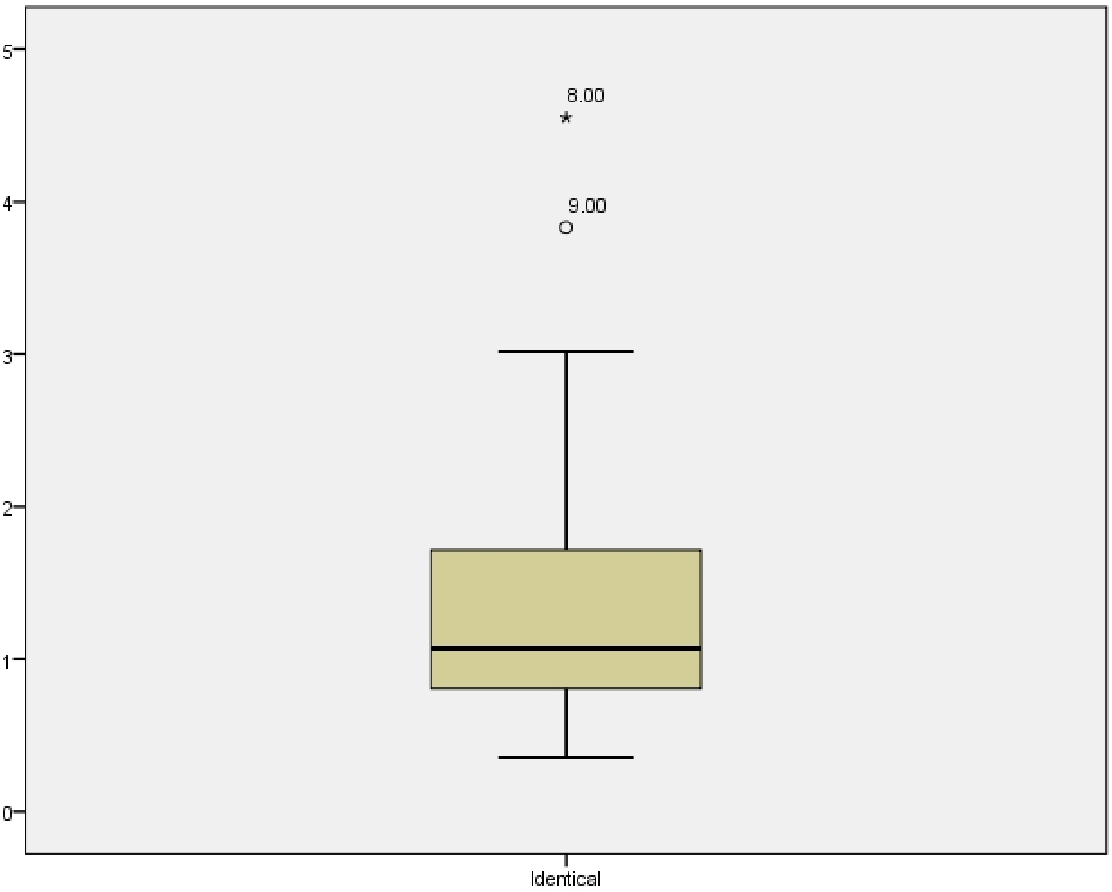
Appendix-5.12: RQ2.1 Example of AR Tool’s Data Output



	A	B	C	D	E	F	G	H	I	J
1	ID	Order	Icon	ViewingTime	aligned(ViewingTime) Order	aligned(ViewingTime) Icon	aligned(ViewingTime) Order*Icon	ART(ViewingTime) Order	ART(ViewingTime) Icon	ART(ViewingTime) Order*Icon
2	1	1	1	1.09	-0.342708333	-3.250416667	-0.360208333	52	17	54
3	1	1	2	2.21	-0.926041667	-1.761666667	-1.312291667	35	39	28
4	1	1	3	2.71	-2.075208333	-1.87	-1.853125	15	37	19
5	1	1	4	4.49	-3.656041667	-0.08375	-3.440208333	4	57	7
6	2	1	1	2.65	1.217291667	-1.690416667	1.199791667	72	40	71
7	2	1	2	6.63	3.493958333	2.658333333	3.107708333	89	79	88
8	2	1	3	9.78	4.994791667	5.2	5.216875	90	88	91
9	2	1	4	13.63	5.483958333	9.05625	5.699791667	93	94	94

Appendix-5.12: An example of AR Tool’s output for data from RQ2.1. Columns A-D were input by the researcher and AR Tool returned columns E-J.

Appendix-5.13: RQ2.2: Outliers Boxplot



Appendix-5.13: Boxplot for RQ2.2 showing the speed of recognition of the participants for the identical icons. Participant 8 is a significant outlier and participant 9 is a nearly significant outlier, as indicated by the z-scores.

Appendix-5.14: RQ2.2 Shapiro-Wilk Normality Test

Tests of Normality

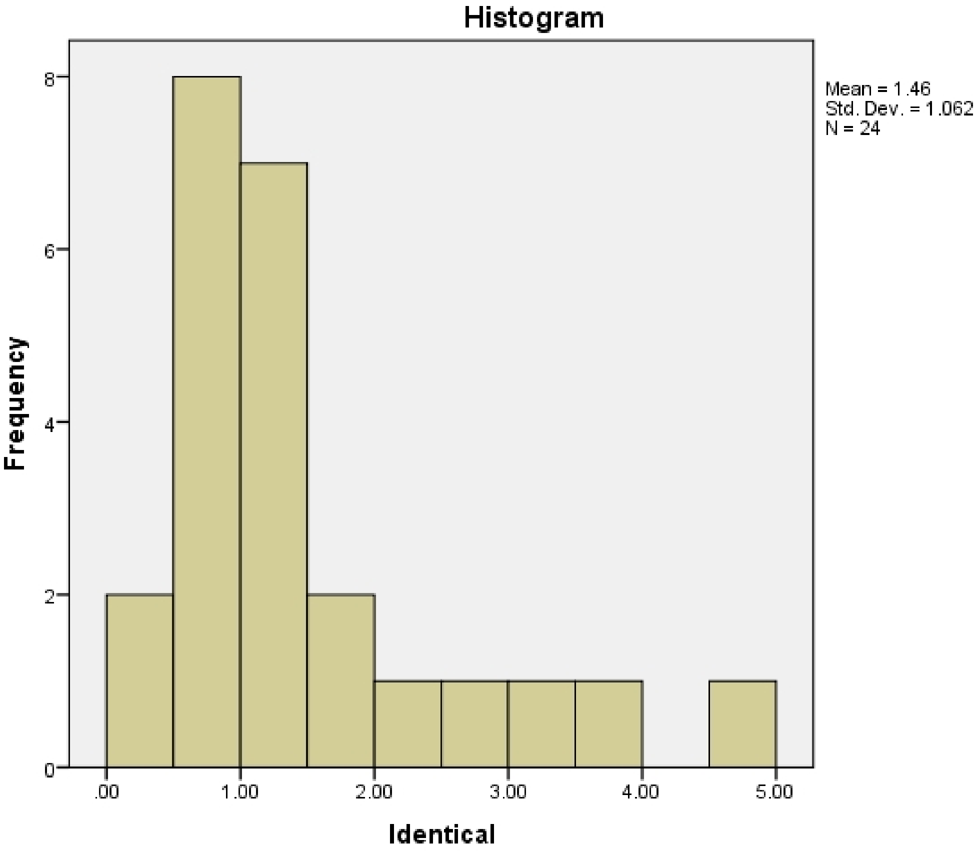
Speed of recognition means	**Shapiro-Wilk**		
	Statistic	**df**	**Sig.**
Identical	.80	24	.000
Slightly different	.91	24	.042
Mostly different	.93	24	.084
Dissimilar	.97	24	.615

*. This is a lower bound of the true significance.

a. Lilliefors significance correction

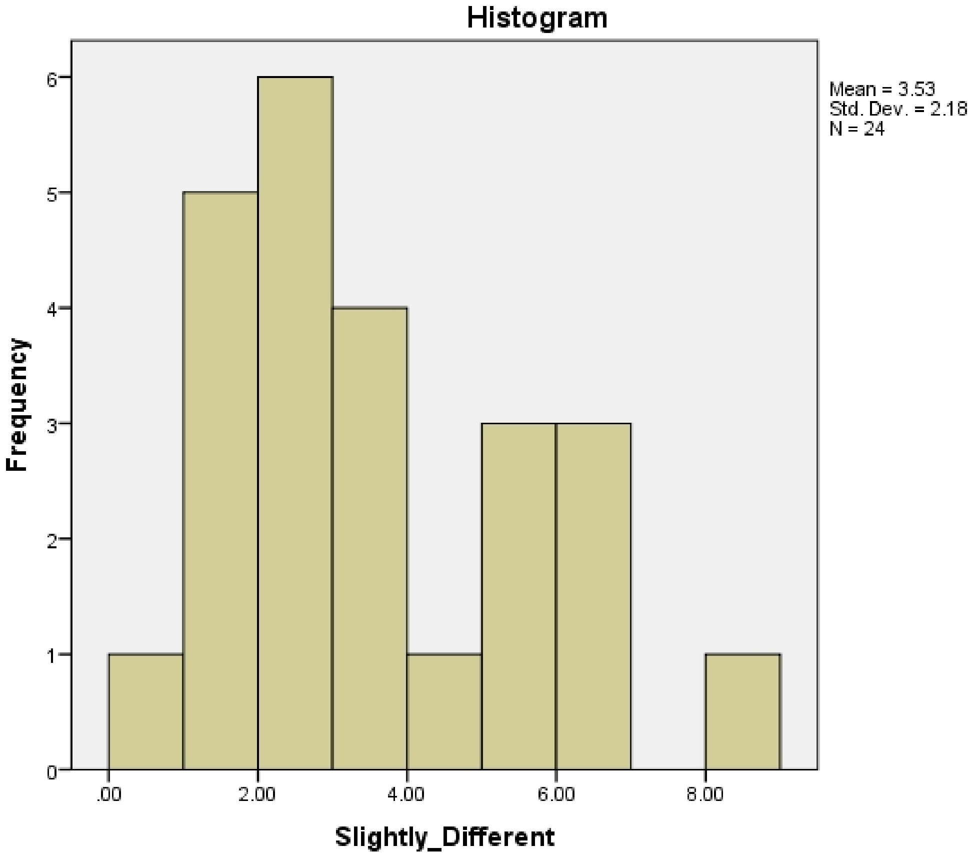
Appendix-5.14: The Shapiro-Wilk test of normality for RQ2.2. The identical and slightly different icons have a p-value below 0.05 (in red), which indicates that they are not normally distributed.

Appendix-5.15: RQ2.2 Histogram: Identical Icons



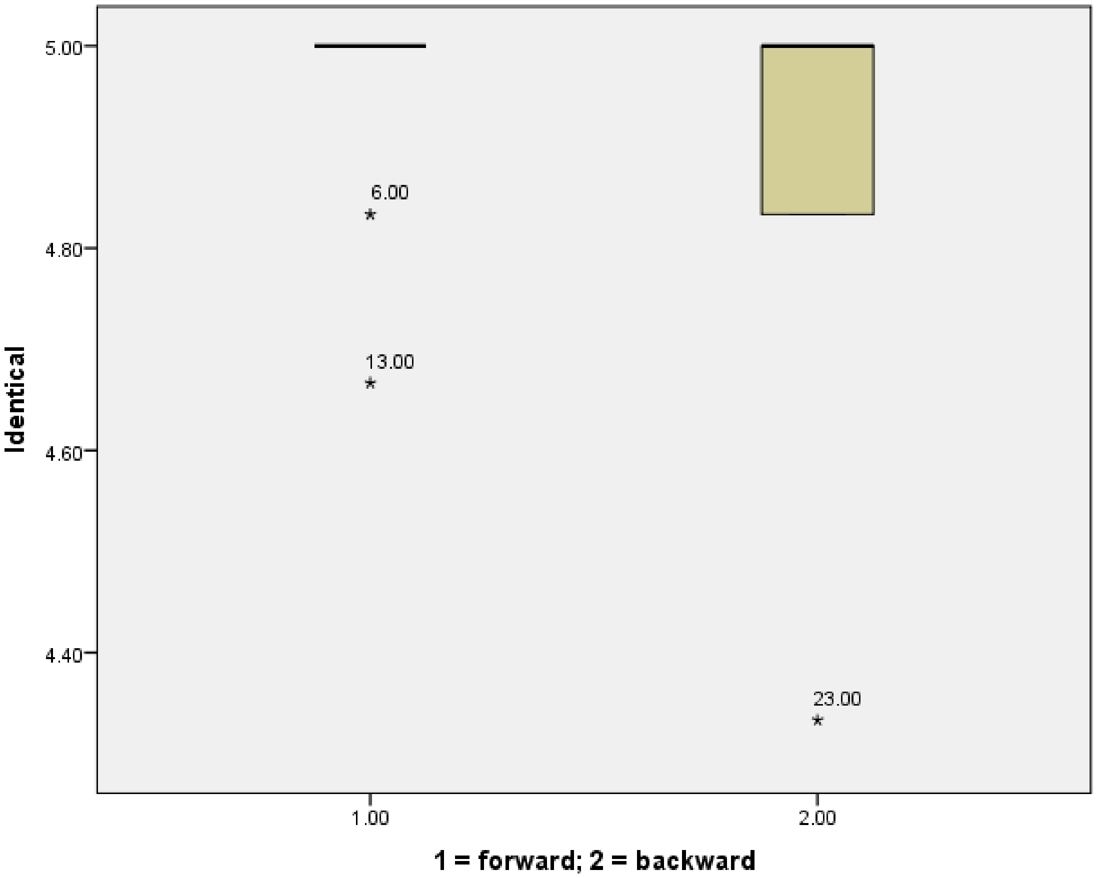
Appendix-5.15: Histogram for RQ2.2 showing the distribution of the participants' speed of recognition times for the identical icons.

Appendix-5.16: RQ2.2 Histogram: Slightly Different



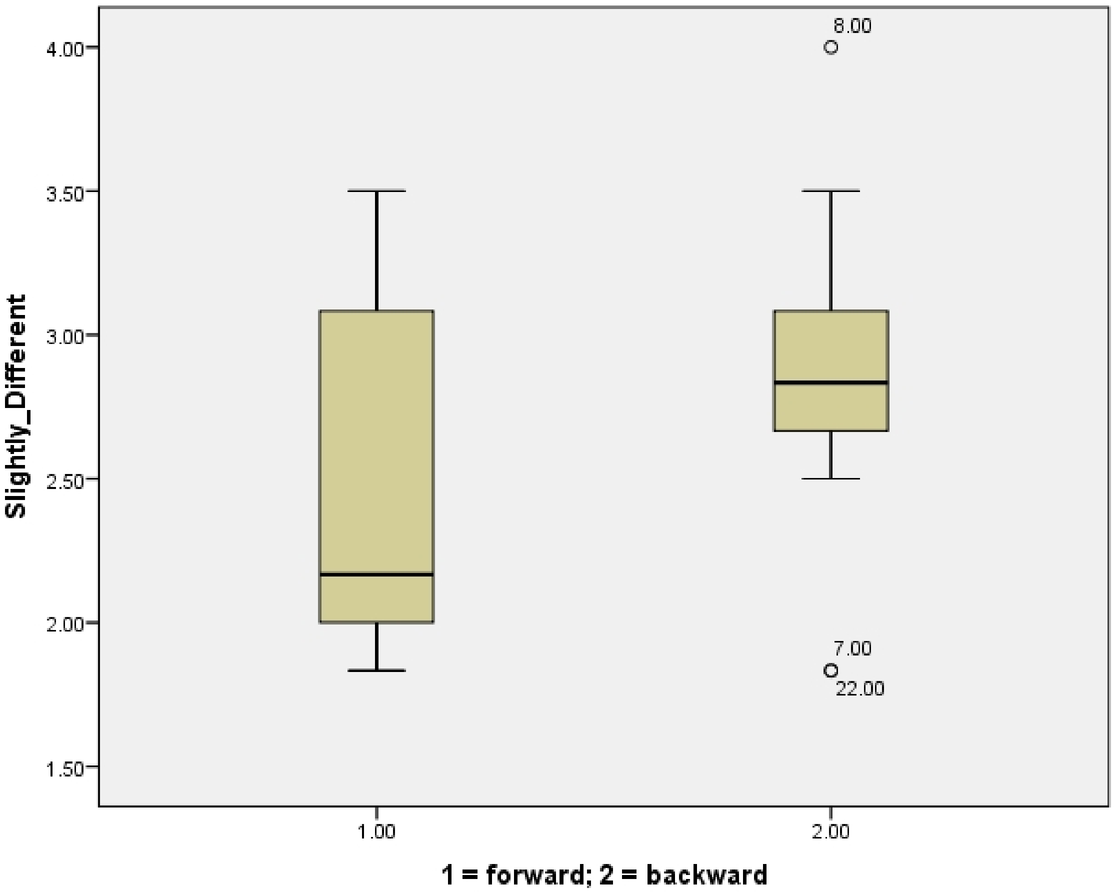
Appendix-5.16: Histogram for RQ2.2 showing the distribution of the participants' speed of recognition times for the slightly different icons.

Appendix-5.17: RQ2.3 Outliers Boxplot: ‘forward’ and ‘backward’ Presentation Orders, Identical Icons



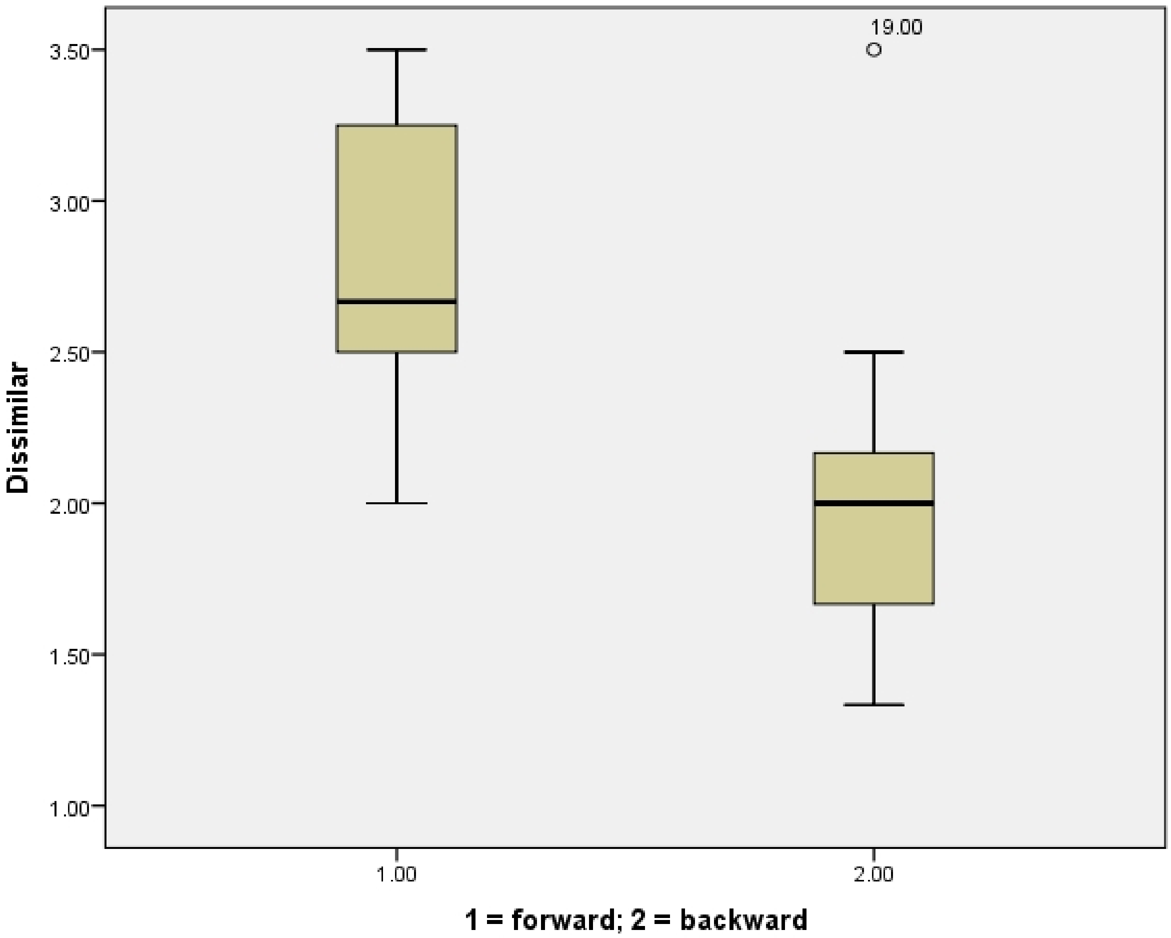
Appendix-5.17: Boxplot for RQ2.3 showing the accuracy of interpretation scores of the participants in the ‘forward’ (left) and ‘backward’ (right) presentation orders, identical icon type conditions. In the ‘forward’ presentation order, identical icon type condition participants 6 and 13 are nearly significant outliers, as indicated by the z-scores. In the ‘backward’ presentation order, identical icon type condition participant 23 is a significant outlier, as indicated by their z-score.

Appendix-5.18: RQ2.3 Outliers Boxplot: ‘backward’ presentation, Slightly Different Icons



Appendix-5.18: Boxplot for RQ2.3 showing the accuracy of interpretation scores of the participants in the ‘forward’ (left) and ‘backward’ (right) presentation orders, slightly different icon type conditions. In the ‘backward’ presentation order, slightly different icon type condition participants 7, 8, 22 are nearly significant outliers, as indicated by their z-scores.

Appendix-5.19: RQ2.3 Outliers Boxplot: ‘backward’ Presentation, Dissimilar Icons



Appendix-5.19: Boxplot for RQ2.3 showing the accuracy of interpretation scores of the participants in the ‘forward’ (left) and ‘backward’ (right) presentation orders, dissimilar icon type conditions. In the ‘backward’ presentation order, dissimilar icon type condition participant 19 is a nearly significant outlier, as indicated by their z-score.

Appendix-5.20: RQ2.3 Shapiro-Wilk Normality Test

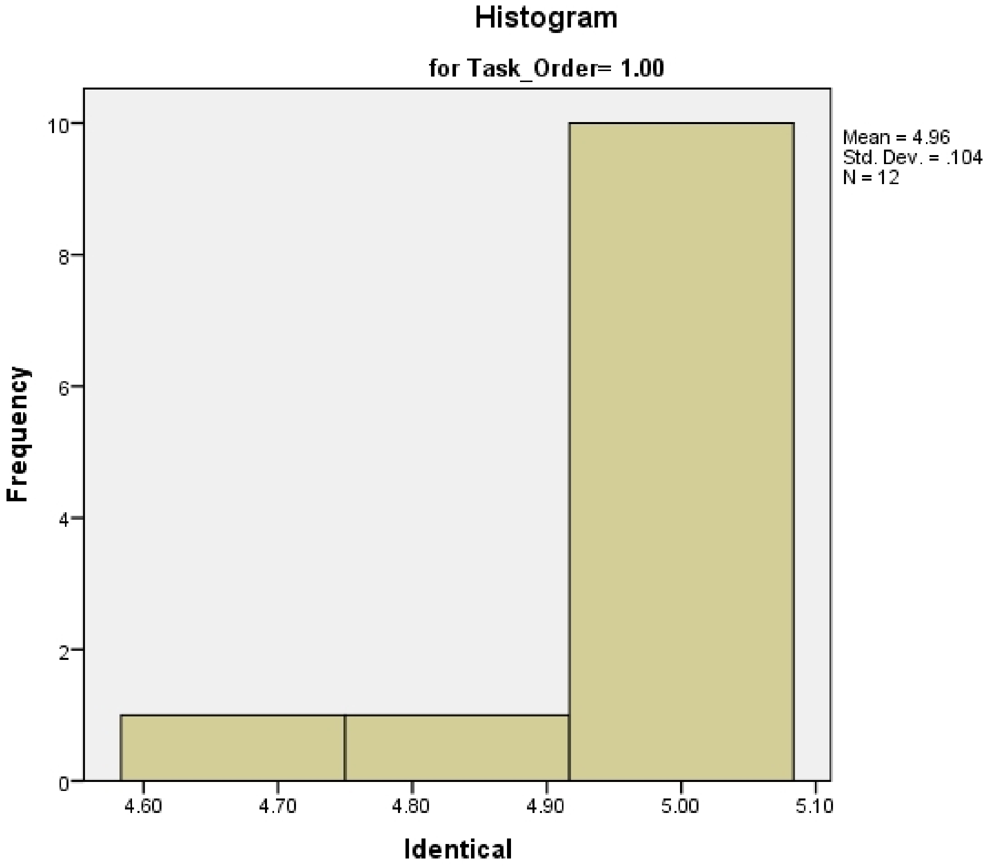
Tests of Normality				
Icon type	**Presentation order**	**Shapiro-Wilk**		
		Statistic	**df**	**Sig.**
Identical	Forward	.479	12	< .001
	Backward	.570	12	< .001
Slightly different	Forward	.830	12	.021
	Backward	.921	12	.294
Mostly different	Forward	.913	12	.232
	Backward	.952	12	.669
Dissimilar	Forward	.905	12	.182
	Backward	.858	12	.046

*. This is a lower bound of the true significance.

a. Lilliefors significance correction

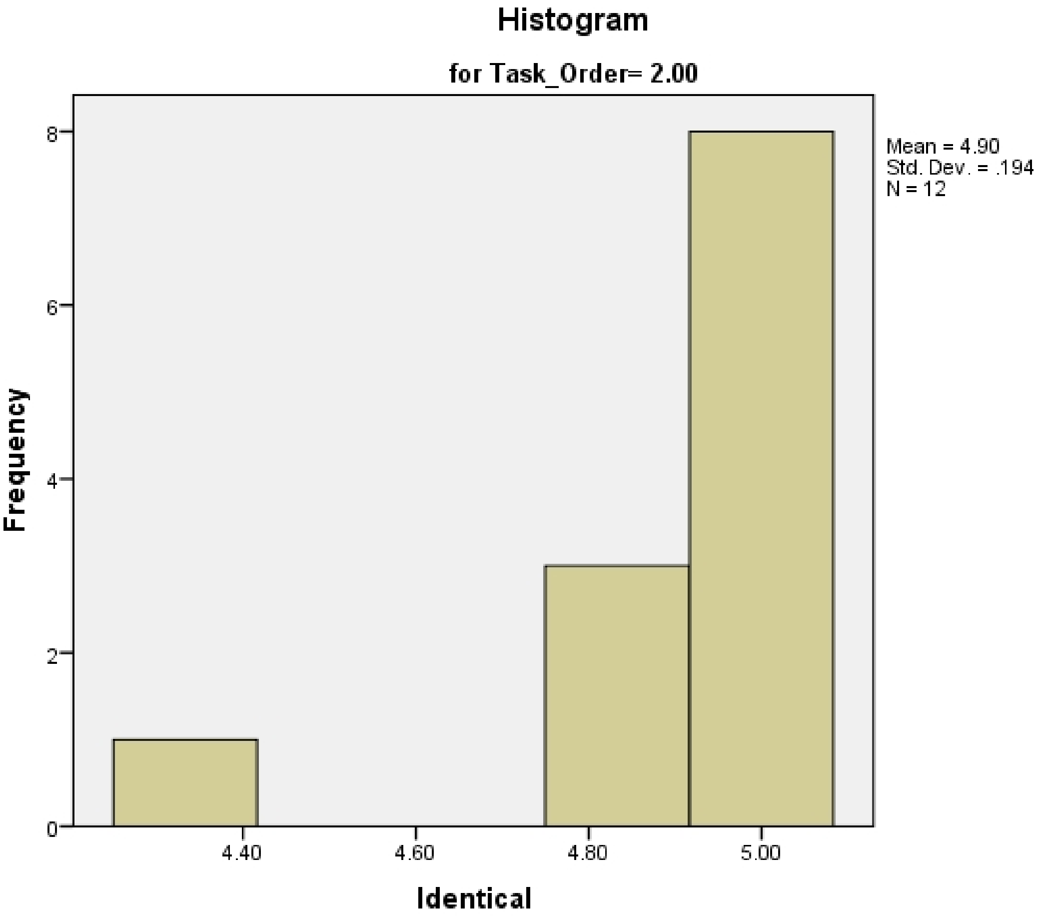
Appendix-5.20: The Shapiro-Wilk test of normality for RQ2.3. The ‘forward’ and ‘backward’ presentation order, identical icon type conditions; the ‘forward’ presentation order, slightly different icon type condition; and the ‘backward’ presentation order, dissimilar icon type condition have p-values below 0.05 (in red), which indicates that they are not normally distributed.

Appendix-5.21: RQ2.3 Histogram: 'forward' Presentation, Identical Icons



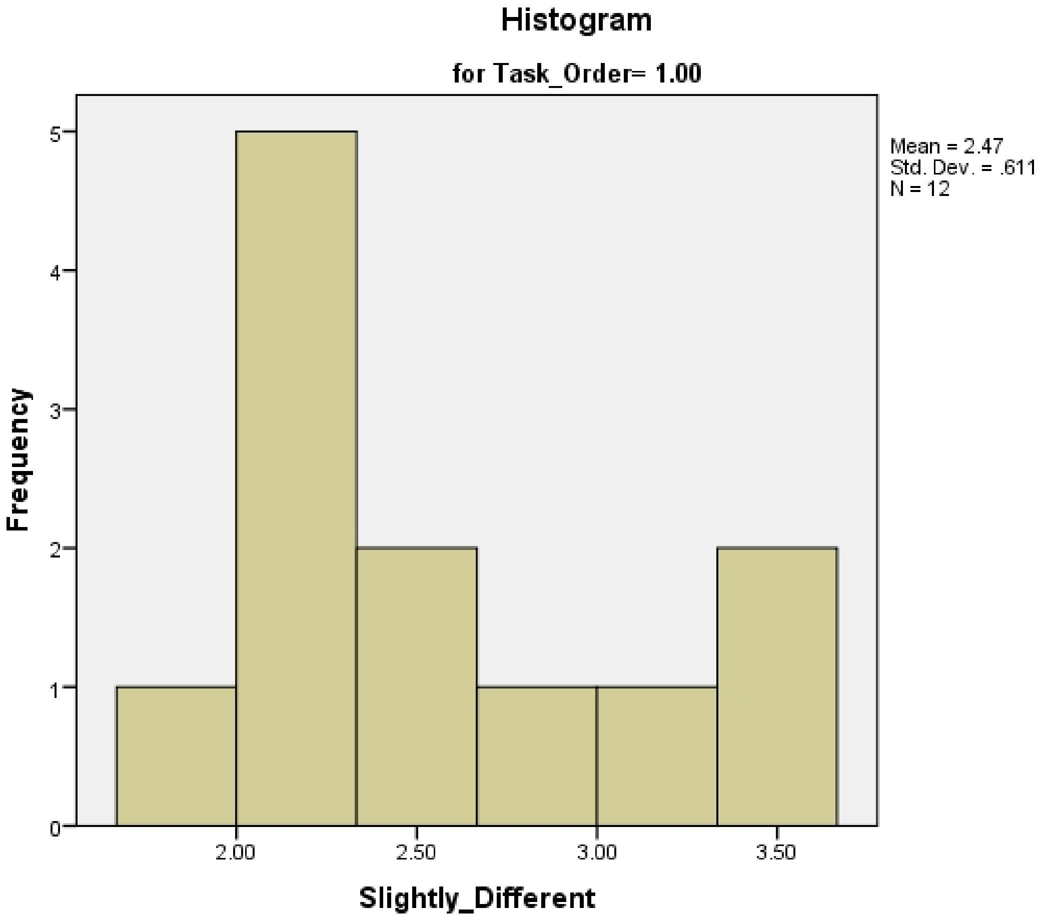
Appendix-5.21: Histogram for RQ2.3 showing the distribution of the participants' accuracy of interpretation scores in the 'forward' presentation order, identical icon type condition.

Appendix-5.22: RQ2.3 Histogram: ‘backward’ Presentation, Identical Icons



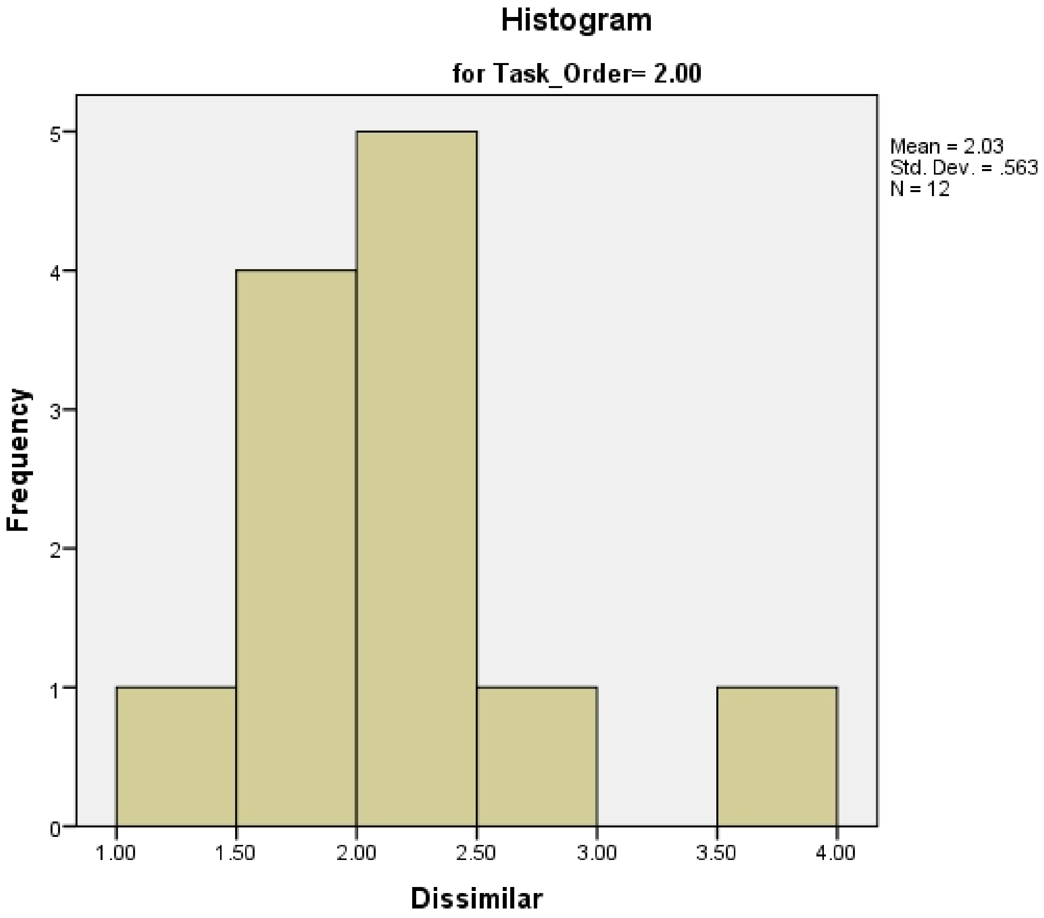
Appendix-5.22: Histogram for RQ2.3 showing the distribution of the participants’ accuracy of interpretation scores in the ‘backward’ presentation order, identical icon type condition.

Appendix-5.23: RQ2.3 Histogram: ‘forward’ Presentation, Slightly Different Icons



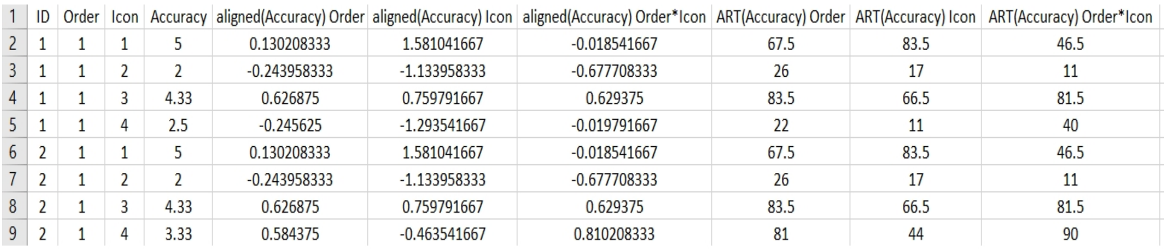
Appendix-5.23: Histogram for RQ2.3 showing the distribution of the participants’ accuracy of interpretation scores in the ‘forward’ presentation order, slightly different icon type condition.

Appendix-5.24: RQ2.3 Histogram: ‘backward’ Presentation, Dissimilar Icons



Appendix-5.24: Histogram for RQ2.3 showing the distribution of the participants’ accuracy of interpretation scores in the ‘backward’ presentation order, dissimilar icon type condition.

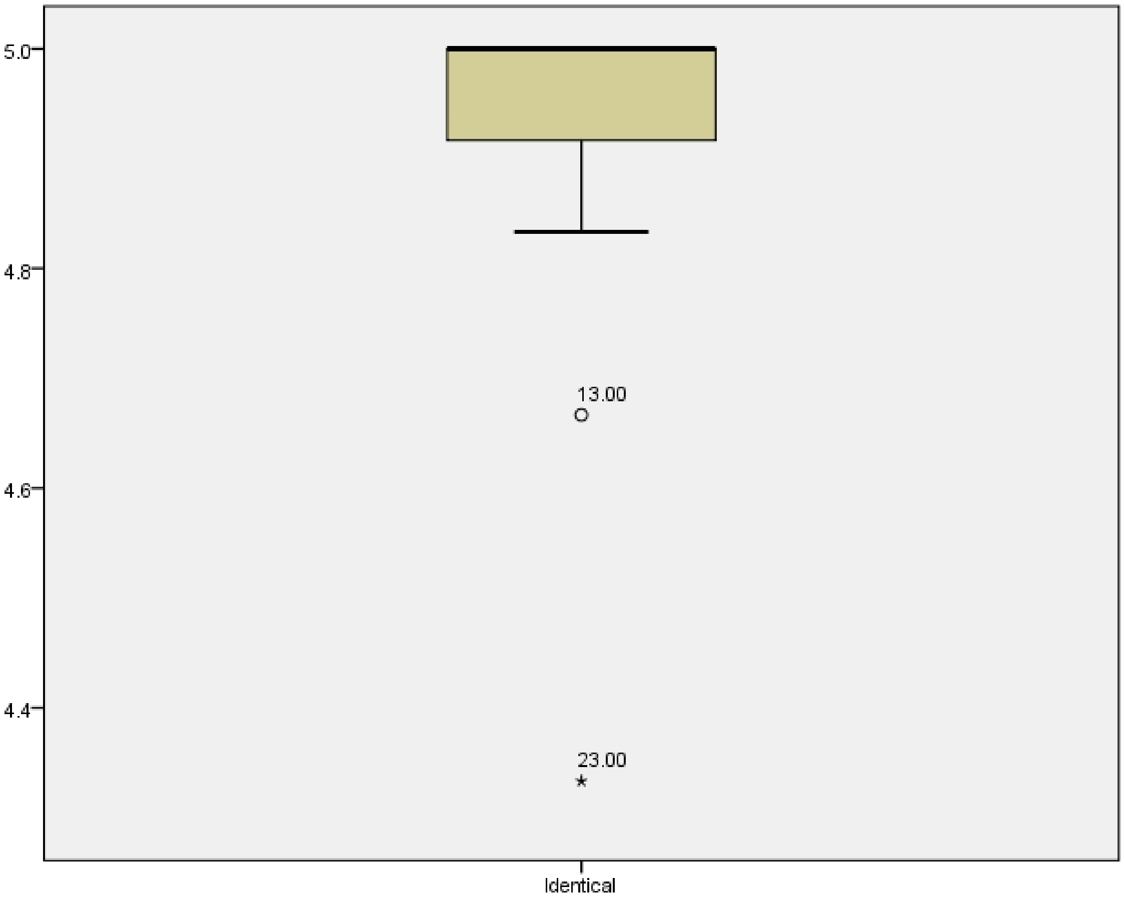
Appendix-5.25: RQ2.3 Example of AR Tool’s Data Output



1	ID	Order	Icon	Accuracy	aligned(Accuracy) Order	aligned(Accuracy) Icon	aligned(Accuracy) Order*Icon	ART(Accuracy) Order	ART(Accuracy) Icon	ART(Accuracy) Order*Icon
2	1	1	1	5	0.130208333	1.581041667	-0.018541667	67.5	83.5	46.5
3	1	1	2	2	-0.243958333	-1.133958333	-0.677708333	26	17	11
4	1	1	3	4.33	0.626875	0.759791667	0.629375	83.5	66.5	81.5
5	1	1	4	2.5	-0.245625	-1.293541667	-0.019791667	22	11	40
6	2	1	1	5	0.130208333	1.581041667	-0.018541667	67.5	83.5	46.5
7	2	1	2	2	-0.243958333	-1.133958333	-0.677708333	26	17	11
8	2	1	3	4.33	0.626875	0.759791667	0.629375	83.5	66.5	81.5
9	2	1	4	3.33	0.584375	-0.463541667	0.810208333	81	44	90

Appendix-5.25: An example of AR Tool’s output for data from RQ2.3. Columns A-D were input by the researcher and AR Tool returned columns E-J.

Appendix-5.26: RQ2.4 Outliers Boxplot



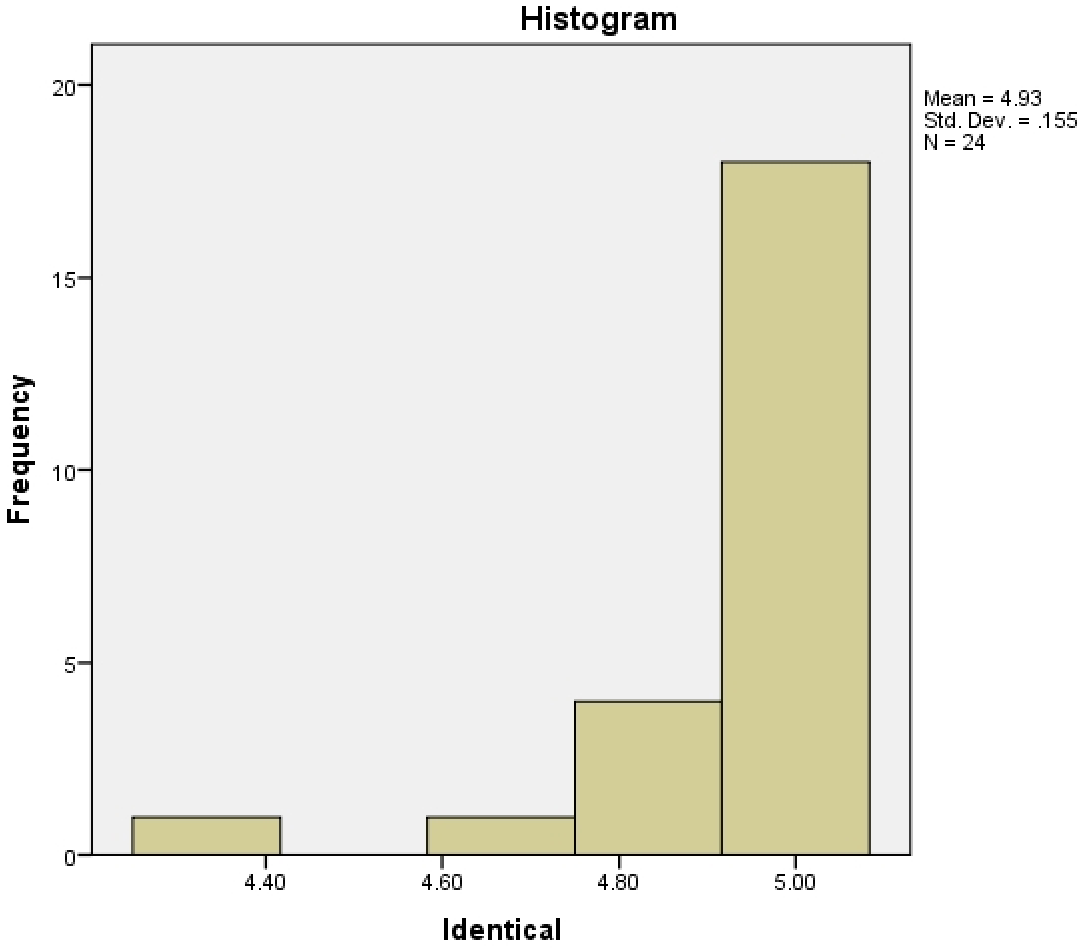
Appendix-5.26: Boxplot for RQ2.4 showing the accuracy of interpretation of the participants for the identical icons. Participant 23 is a significant outlier and participant 13 is a nearly significant outlier, as indicated by the z-scores.

Appendix-5.27: RQ2.4 Shapiro-Wilk Normality Test

Tests of Normality			
Speed of recognition means	Shapiro-Wilk		
	Statistic	df	Sig.
Identical	.52	24	.000
Slightly different	.93	24	.084
Mostly different	.94	24	.173
Dissimilar	.94	24	.170
*. This is a lower bound of the true significance.			
a. Lilliefors significance correction			

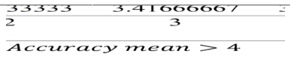
Appendix-5.27: The Shapiro-Wilk test of normality for RQ2.4. The identical icons have a p-value below 0.05 (in red), which indicates that they are not normally distributed.

Appendix-5.28: RQ2.4 Histogram



Appendix-5.28: Histogram for RQ2.4 showing the distribution of the participants' accuracy of interpretation for the identical icons.

Appendix-6.1: Slightly Different Icons from Study II and their Accuracy Means

			Slightly different			
	User Account	Email	Password	Review	Like	Map
		Flight			Hotel	
			Forward			
Observed	2.16666667	1.91666667	1.83333333	3	3	2.91666667
Rounded values	2	2	2	3	3	3
			Backward			
Observed	2.66666667	2.58333333	2.33333333	3.41666667	3.91666667	2.16666667
Rounded values	3	3	2	3	4	2
Selection Range			3 >  33333 3.41666667 2 3 *Accuracy mean > 4*			

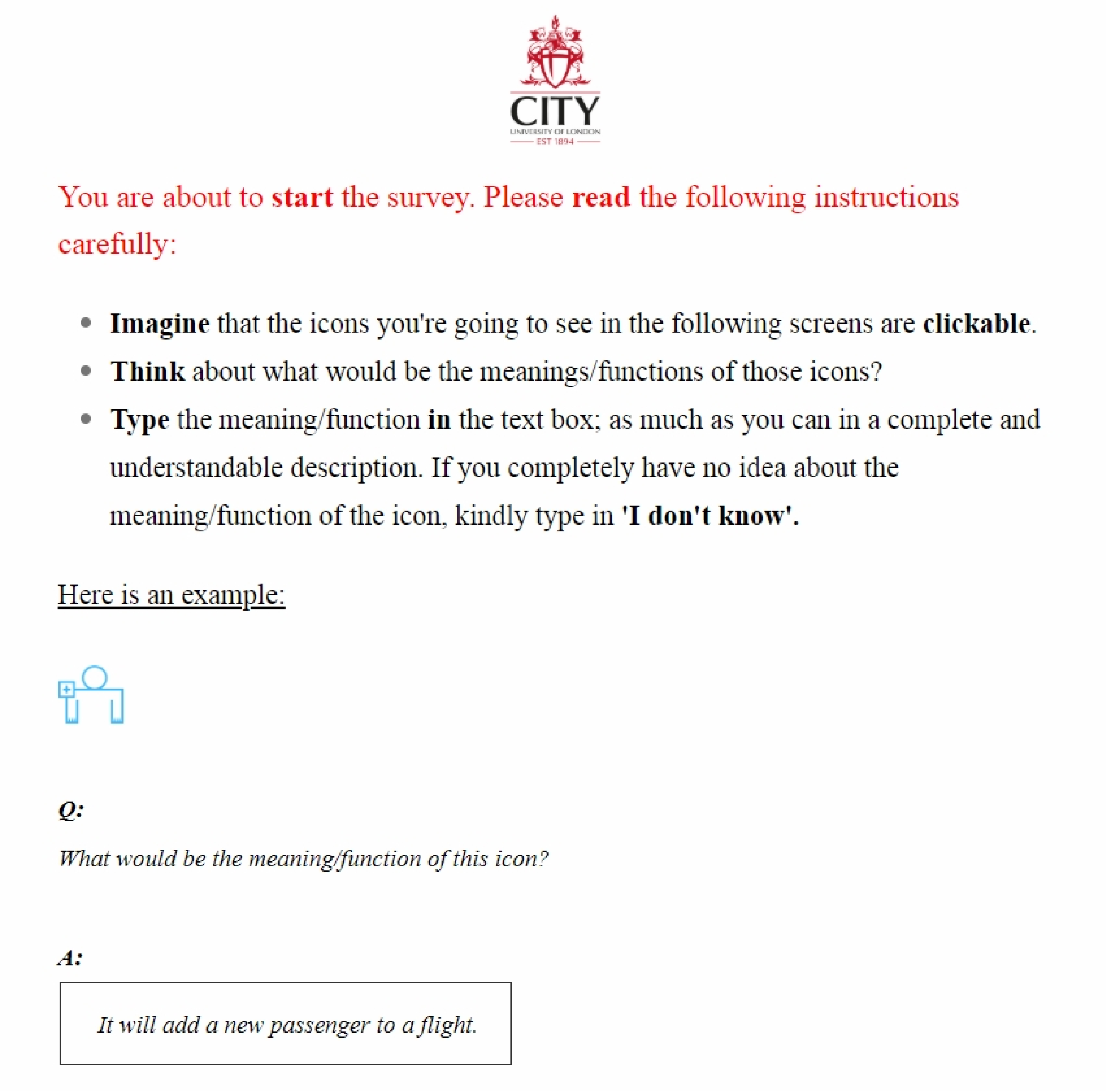
Appendix-6.2: Mostly Different Icons from Study II and their Accuracy Means

	Mostly different					
	User Account	Email	Password	Review	Like	Map
	Flight			Hotel		
Observed	3.50	2.88	2.71	3.67	4.38	4.88
Rounded values	4	3	3	4	4	5
Selection Range			2 > 3.67 4 $Accuracy\ mean > 3$			

Appendix-6.3: Completely Different Icons from Study II and their Accuracy Means

			Completely different			
	User Account	Email	Password	Review	Like	Map
	Flight			Hotel		
			Forward			
Observed	3.83333333	2.25	2.25	3.16666667	2.58333333	2.58333333
Rounded values	4	2	2	3	3	3
			Backward			
Observed	1.83333333	1.83333333	1.66666667	2.16666667	2.25	2.41666667
Rounded values	2	2	2	2	2	2
Selection Range			1 > 2 2 *Accuracy mean* > 2			

Appendix-6.4: Instructional Page of the ‘No Visual Context’ Group



You are about to **start** the survey. Please **read** the following instructions carefully:

- **Imagine** that the icons you're going to see in the following screens are **clickable**.
- **Think** about what would be the meanings/functions of those icons?
- **Type** the meaning/function **in** the text box; as much as you can in a complete and understandable description. If you completely have no idea about the meaning/function of the icon, kindly type in **'I don't know'.**

Here is an example:

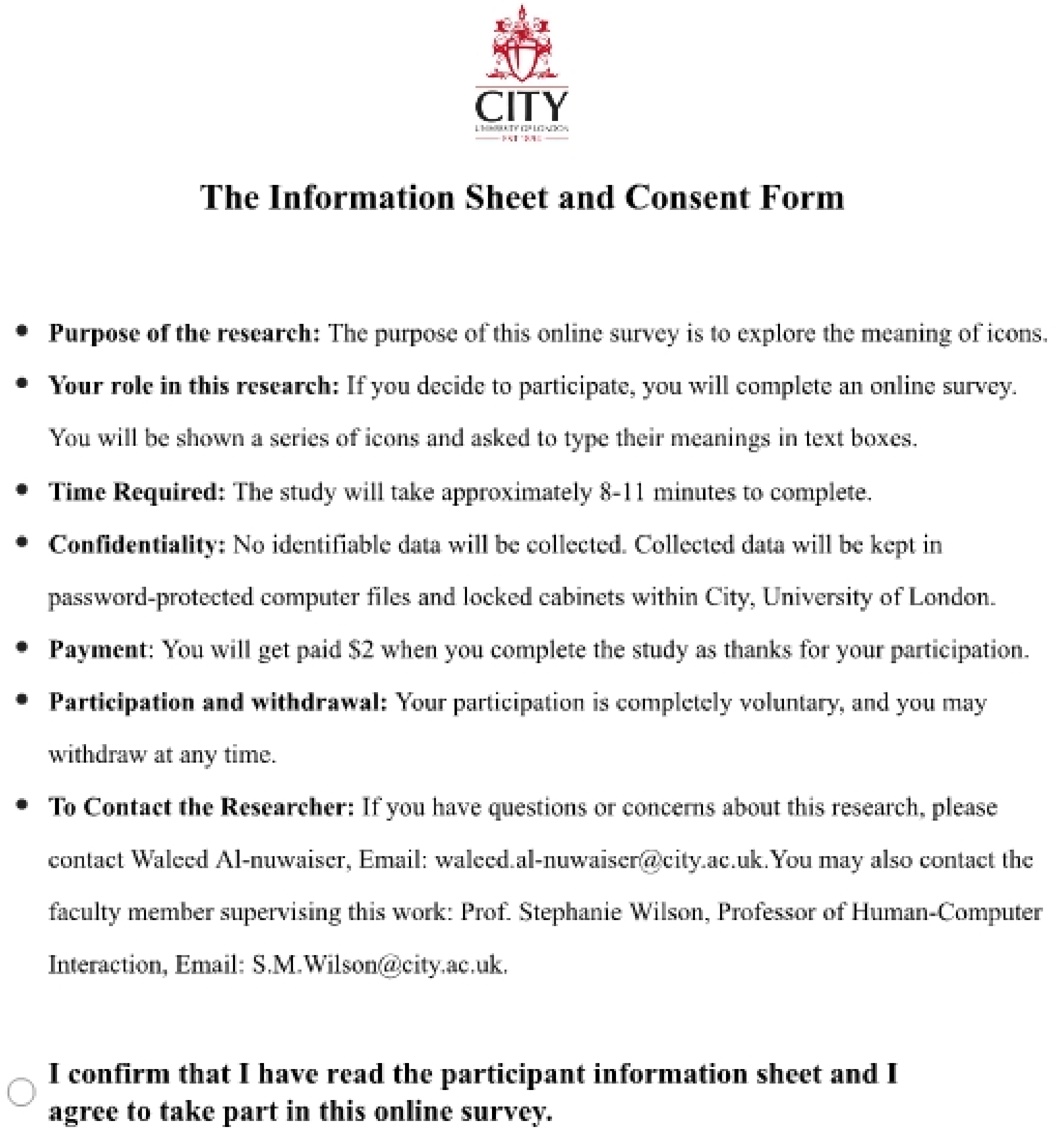
Q:

What would be the meaning/function of this icon?

A:

It will add a new passenger to a flight.

Appendix-6.5: Online Consent Form

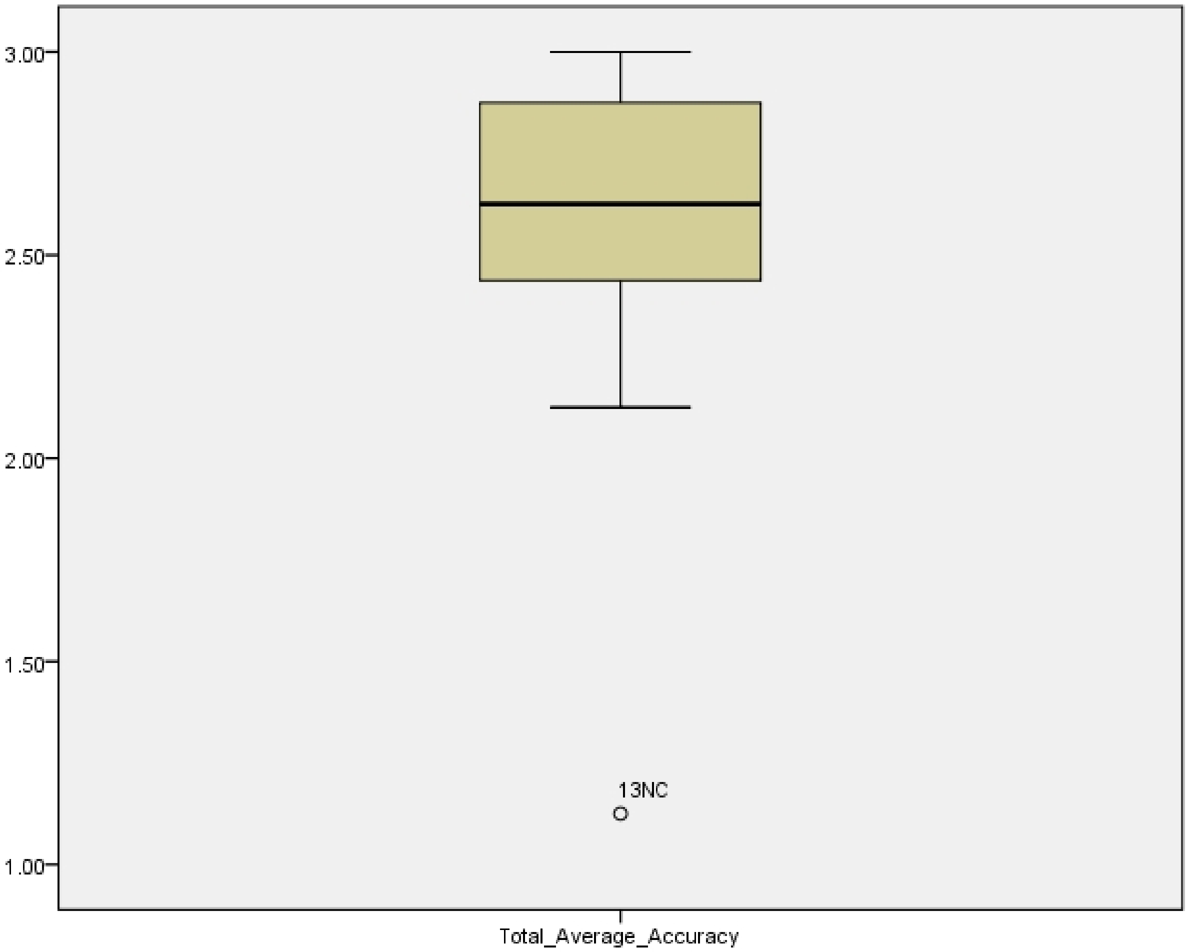


The Information Sheet and Consent Form

- **Purpose of the research:** The purpose of this online survey is to explore the meaning of icons.
- **Your role in this research:** If you decide to participate, you will complete an online survey. You will be shown a series of icons and asked to type their meanings in text boxes.
- **Time Required:** The study will take approximately 8-11 minutes to complete.
- **Confidentiality:** No identifiable data will be collected. Collected data will be kept in password-protected computer files and locked cabinets within City, University of London.
- **Payment**: You will get paid $2 when you complete the study as thanks for your participation.
- **Participation and withdrawal:** Your participation is completely voluntary, and you may withdraw at any time.
- **To Contact the Researcher:** If you have questions or concerns about this research, please contact Waleed Al-nuwaiser, Email: waleed.al-nuwaiser@city.ac.uk.You may also contact the faculty member supervising this work: Prof. Stephanie Wilson, Professor of Human-Computer Interaction, Email: S.M.Wilson@city.ac.uk.

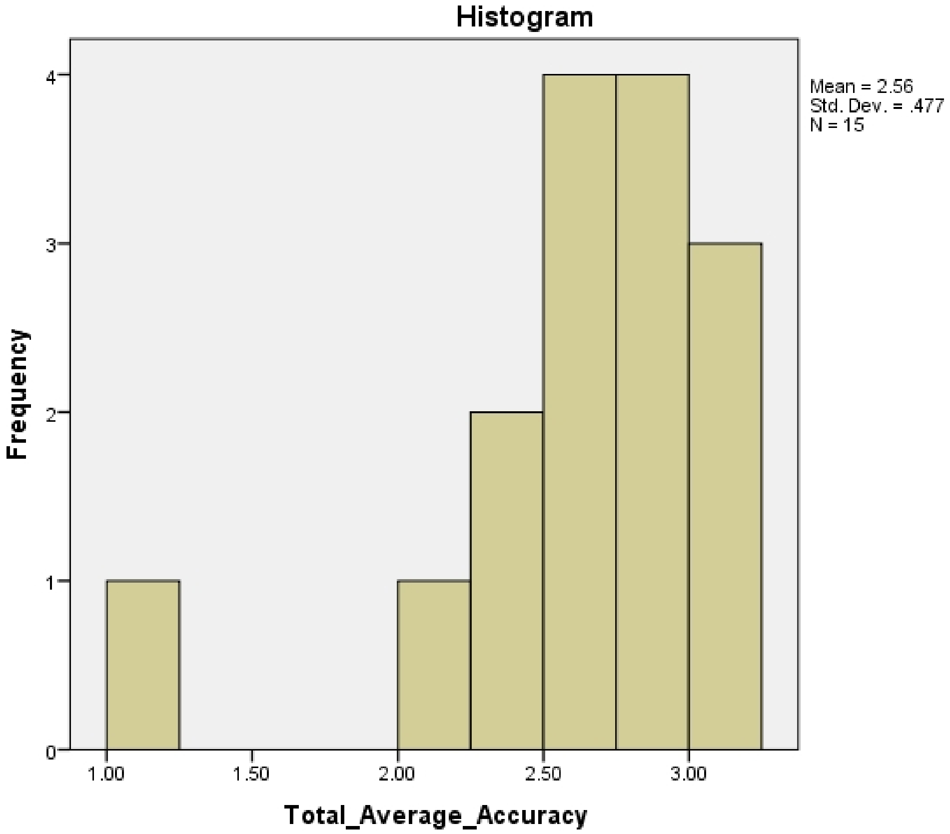
○ **I confirm that I have read the participant information sheet and I agree to take part in this online survey.**

Appendix-6.6: RQ3: Outliers Boxplot



Appendix-6.6: Boxplot for RQ3 showing the accuracy of interpretation scores of the participants in the ‘no visual context’ group. Participant 13NC is a significant outlier, as indicated by the z-scores.

Appendix-6.7: RQ3 Histogram



Appendix-6.7: Histogram for RQ3 showing the distribution of the participants' accuracy of interpretation scores in the no-visual-context group.

Chapter 9 Bibliography

Adamson, P.J., 1996. A comparison between consistent and inconsistent graphical user interfaces (Doctoral dissertation, University of North Florida).

Alexander, R.G. and Zelinsky, G.J., 2011. Visual similarity effects in categorical search. Journal of Vision, 11(8), pp.9-9.

Ali, A., Alrasheedi, M., Ouda, A. And Capretz, L. F., 2015. A study of the interface usability issues of mobile learning applications for smartphones from the user's perspective. [online] Available at: <https://arxiv.org/abs/1501.01875> [Accessed 20 August 2016].

Alleydog's Online Glossary., 2019. Psychology Glossary | alleydog.com. [online] Available at: http://www.alleydog.com/glossary [Accessed 15 May. 2019].

Ark, W., Dryer, D.C., Selker, T. and Zhai, S., 1998. Representation matters: The effect of 3D objects and a spatial metaphor in a graphical user interface. In People and Computers XIII (pp. 209-219). Springer, London.

Atkinson, R. C. And Shiffrin, R. M., 1968. Human memory: A proposed system and its control processes. In: SPENCE, K. W. And SPENCE, J. T., 1968. The psychology of learning and motivation (Volume 2). New York: Academic Press. Pp. 89–195.

Baddeley, A., 2000. The episodic buffer: a new component of working memory? Trends in Cognitive Sciences, 4(11), pp. 417-423.

Baddeley, A.D., 2014. Essentials of human memory. Hove, UK: Psychology Press Classic Editions.

Bahrick, H.P. and Boucher, B., 1968. Retention of visual and verbal codes of the same stimuli. Journal of Experimental Psychology, 78(3p1), p.417.

Banerjee, S. and Pedersen, T., 2002, February. An adapted Lesk algorithm for word sense disambiguation using WordNet. In International conference on intelligent text processing and computational linguistics (pp. 136-145). Springer, Berlin, Heidelberg.

Bedford, A., 2014. Icon usability. Nngroup Usability Report. [online] Available at: <http://www. Nngroup. Com/articles/icon-usability/> [Accessed 20 August 2016].

Ben-Av, M.B. And Sagi D., 1995. Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. Vision Research, 35, pp.853–866.

Bergstrom, J.R. and Schall, A. eds., 2014. Eye-tracking in user experience design. Elsevier.

Bermeitinger, C., 2014. Priming. 10.4018/978-1-4666-6599-6.ch002.

Bhaskar, R., 2013. A realist theory of science. Routledge.

Blankenberger, S. and Hahn, K., 1991. Effects of icon design on human-computer interaction. International Journal of Man-Machine Studies, 35(3), pp.363-377.

Boeriis, M. and Holsanova, J., 2012. Tracking visual segmentation: connecting semiotic and cognitive perspectives. Visual communication, 11(3), pp.259-281.

Bogdani, August 25, 2016. Compute sentence similarity using Wordnet. [online] Available at: < https://nlpforhackers.io/author/bogdani/page/3/>

Borji, A., and Itti, L. 2015. Cat2000: A large scale fixation dataset for boosting saliency research. arXiv preprint arXiv:1505.03581.

Borkin, M.A., Bylinskii, Z., Kim, N.W., Bainbridge, C.M., Yeh, C.S., Borkin, D., Pfister, H. and Oliva, A., 2016. Beyond memorability: Visualization recognition and recall. IEEE transactions on visualization and computer graphics, 22(1), pp.519-528.

Bouissac, P., 1999. Semiotics and the science of memory. In: Conference on Semiotics and the European Heritage. Dresden, February 1999.

Brady, T.F., Konkle, T. and Alvarez, G.A., 2011. A review of visual memory capacity: Beyond individual items and toward structured representations. Journal of vision, 11(5), pp.4-4.

Braun, V. and Clarke, V., 2013. Successful qualitative research: A practical guide for beginners. sage.

Budanitsky, A. and Hirst, G., 2006. Evaluating wordnet-based measures of lexical semantic relatedness. Computational Linguistics, 32(1), pp.13-47.

Budiu, Raluca, 2014, Memory Recognition and Recall in User Interfaces, viewed 1 April 2019, < https://www.nngroup.com/articles/recognition-and-recall/>

Bugelski, B.R., 1962. Presentation time, total time, and mediation in paired-associate learning. Journal of Experimental Psychology, 63(4), p.409.

Burgess, N. And Hitch, G., 2005. Computational models of working memory: Putting long-term memory into context. Trends in Cognitive Sciences, 9(11), pp.535-541.

Burigat, S. and Chittaro, L. 2011, Visualizing references to off-screen content on mobile devices: A comparison of Arrows, Wedge, and Overview + Detail. Interacting with Computers, 23(2), pp.156-166.

Buswell, G.T., 1935. How people look at pictures: A study of the psychology of perception in art.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A. And Oliva, A., 2015. Intrinsic and extrinsic effects on image memorability. Vision research, 116, pp.165-178.

Bzostek, J. A., and Wogalter, M. S., 1999. Measuring visual search time for a product warning label as a function of icon, color, column and vertical placement. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 43(16), 888–892. https://doi.org/10.1177/154193129904301603.

Carlei, C. and Kerzel, D., 2014. Gaze direction affects visuo-spatial short-term memory. Brain and cognition, 90, pp.63-68.

Charles Zaiontz, 2012. Scheirer-Ray-Hare Test, Real Statistics Using Excel. [online] Available at: http://www.real-statistics.com/two-way-anova/scheirer-ray-hare-test/

Cheng, H.I. and Patterson, P.E., 2007. Iconic hyperlinks on e-commerce websites. Applied Ergonomics, 38(1), pp.65-69.

Conover, W.J. and Iman, R.L., 1981. Rank transformations as a bridge between parametric and non-parametric statistics. The American Statistician, 35(3), pp.124-129.

Curtis, S.A., 2003. The classification of greedy algorithms. Science of Computer Programming, 49(1-3), pp.125-157.

De Angeli, A., Sutcliffe, A. and Hartmann, J., 2006, June. Interaction, usability and aesthetics: what influences users' preferences? In Proceedings of the 6th conference on Designing Interactive systems (pp. 271-280). ACM. devi: A comparison of Arrows, Wedge, and Overview + Detail. Interacting with devices doi:10.3758/BF03196470.

De S Reis, S. And Prates, R.O., 2012. Assessing the semiotic inspection method: the evaluators' perspective. In: Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems, pp. 287-296. Brazilian Computer Society.

De Saussure, F., 1916. Course in general linguistics. London: Duckworth.

De Souza, C. S. And Leitão, C. F., 2009. Semiotic engineering methods for scientific research in HCI. Synthesis Lectures on Human-Centered Informatics, 2, pp.1-122.

De Souza, C. S., Leitão, C. F., Prates, R. O. And Da Silva, E. J. E., 2006. The semiotic inspection method. In: Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems. New York, NY, USA, ACM.

Delorme, R., 1999, November. Realism in economics: Critical or complex. In Xth Conference of The European Association for Evolutionary Political Economy Charles University of Prague (pp. 4-7).

Derboven, J., De Roeck, D. And Verstraete, M., 2012. Semiotic analysis of multi-touch interface design: The mutable case study. International Journal of Human-Computer Studies, 70, pp.714-728.

Dessus, P. And Peraya, D., 2007. Cognitive modeling of icon comprehension. Enhancing learning through HCI. Hershey: Idea Group Publishing.

Deviant, S., 2011. The Practically Cheating Statistics Handbook--. Lulu. com.

Dey, A. K. And Häkkilä, J., 2008. Context-awareness and mobile devices. User Interface Design and Evaluation for Mobile Technology, 1, pp.205-217.doi:10.1037/0278-7393.26.3.683

Donald, A. N., 2002. The design of everyday things. New York: Basic Books, Inc.

Driver, J., Davis, G., Russell, C., Turatto, M., Freeman, E., 2001. Segmentation, attention, and phenomenal visual objects. Cognition, 80, pp.61–95.

Droll, J.A. and Hayhoe, M.M., 2007. Trade-offs between gaze and working memory use. Journal of Experimental Psychology: Human Perception and Performance, 33(6), p.1352.

Duchnicky, R.L. And Kolers, P.A., 1983. Readability of text scrolled on visual display terminals as a function of window size. Human Factors, 25(6), pp.683-692.

Duncan, J., and Humphreys, G. W. (1989). Visual search and stimulus similarity. Psychological Review, 96(3), 433–458. https://doi.org/10.1037//0033-295x.96.3.433.

Duncan, J., 1984. Selective attention and the organization of visual information. Journal of Experimental Psychology: General, 113(4), p.501.

Eco, U., 1979. A theory of semiotics / Umberto Eco. Bloomington: Indiana University Press.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., and Oliva, A., 2009. Modelling search for people in 900 scenes: A combined source model of eye guidance. Visual cognition, 17(6-7), 945-978.

Estes, W.K., 1972. Interactions of signal and background variables in visual processing. Perception & Psychophysics, 12(3), pp.278-286.

Etikan, I., Musa, S.A. and Alkassim, R.S., 2016. Comparison of convenience sampling and purposive sampling. American journal of theoretical and applied statistics, 5(1), pp.1-4.

Ferreira, J., Barr, P. And Noble, J., 2005. The semiotics of user interface redesign. Proceedings of the Sixth Australasian Conference on User interface, 40. Australian Computer Society.

Fougnie, D., 2008. The relationship between attention and working memory. New research on short-term memory, 1, p.45.

Freeman, D. Benko, H. Ringel, M. and Wigdor, D., 2009. ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. In Procs. ACM Intl. Conf. on Interactive Tabletops and Surfaces (ITS '09). Pages 165-172.

Frost, N., 1972. Encoding and retrieval in visual memory tasks. Journal of Experimental Psychology, 95(2), p.317.

Gatsou, C., Polities, A, and Zevgolis, D. 2012. The importance of mobile interface icons on user interaction. International Journal of Computer Science and Applications. Vol.9, No. 3, pp. 92 – 107.

Gatsou, C., Politis, A. And Zevgolis, D., 2011, September. From icons perception to mobile interaction. In Computer Science and Information Systems (fedcsis), 2011 Federated Conference on (pp. 705-710). IEEE.

Glanzer, M. And Cunitz, A.R., 1966. Two storage mechanisms in free recall. Journal of Verbal Learning and Verbal Behavior, 5(4), pp.351-360.

Goonetilleke, R.S., Shih, H.M. and Fritsch, J., 2001. Effects of training and representational characteristics in icon design. International Journal of Human-Computer Studies, 55(5), pp.741-760.

Green, A. and Barnard, P., 1990. Graphical and iconic interfacing: the role of icon distinctiveness and fixed or variable screen locations.

Griffin, Z., Henderson, J. and Ferreira, F., 2004. The integration of language, vision, and action: eye movements and the visual world. The integration of language, vision, and action: Eye movements and the visual world, pp.00-00.

Guan, Z., Lee, S., Cuddihy, E. and Ramey, J., 2006, April. The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 1253-1262). ACM.

Han, S., Humphreys, G.W. And Chen, L., 1999. Uniform connectedness and classical Gestalt principles of perceptual grouping. Perception & Psychophysics, 61, pp.661–674.

Hancock, P.A., Sawyer, B.D. and Stafford, S., 2015. The effects of display size on performance. Ergonomics, 58(3), pp.337-354.

Haramundanis, K., 1996. Why icons cannot stand alone. ACM SIGDOC Asterisk Journal of Computer Documentation, 20(2), pp.1-8.

Harrison, C., 2003. Visual social semiotics: Understanding how still images make meaning. Technical communication, 50(1), pp.46-60.

Hasher, L. and Zacks, R.T., 1979. Automatic and effortful processes in memory. Journal of experimental psychology: General, 108(3), p.356.

Higgins, J.J. and Tashtoush, S., 1994. An aligned rank transform test for interaction. Nonlinear World, 1(2), pp.201-211.

Hiltunen, K.-M., Hakkila, J. And Tuomela, U., 2005. Subjective understanding of context attributes: a case study. Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future. Computer-Human Interaction Special Interest Group (CHISIG) of Australia.

Holiday Semiotics, 2016. Login page [online] Available at: http://holidaysemiotics.esy.es/Mobile [Accessed 20 September 2016].

Holsanova, J., 2011. How we focus attention in picture viewing, picture description, and during mental imagery (pp. 291-313). von Halem.

Hout, M.C., Godwin, H.J., Fitzsimmons, G., Robbins, A., Menneer, T. and Goldinger, S.D., 2016. Using multidimensional scaling to quantify similarity in visual search and beyond. Attention, Perception, & Psychophysics, 78(1), pp.3-20.

Howell, W.C. and Fuchs, A.H., 1968. Population stereotypy in code design. Organizational Behaviour and Human Performance, 3(3), pp.310-339.

Huang, S.C. and Bias, R.G., 2012. Without context, icons are significantly worse than texts to convey meanings in terms of accuracy and efficiency. ASIST'12.

Huang, S.C., Bias, R.G. and Schnyer, D., 2015. 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), pp.702-720.

Isherwood, S., 2009, July. Graphics and semantics: The relationship between what is seen and what is meant in icon design. In International Conference on Engineering Psychology and Cognitive Ergonomics (pp. 197-205). Springer, Berlin, Heidelberg

Isherwood, S.J., McDougall, S.J. and Curry, M.B., 2007. Icon identification in context: The changing role of icon characteristics with user experience. Human Factors, 49(3), pp.465-476.

Islam, M. N. And Bouwman, H., 2015. An assessment of a semiotic framework for evaluating user-intuitive Web interface signs. Universal Access in the Information Society, 14, pp.563-582.

Islam, M.N. And Bouwman, H., 2016. Towards user–intuitive web interface sign design and evaluation: A semiotic framework. International Journal of Human-Computer Studies, 86, pp.121-137.

Islam, M.N., Tetard F., 2014. Exploring the impact of interface signs' interpretation accuracy, design, and evaluation on web usability: A semiotics perspective, Journal of Systems and Information Technology, Vol. 16 Issue: 4, pp.250-276.

ISO9186-1:2014., 2015. Graphical symbols -- Test methods -- Part 1: Method for testing comprehensibility. [online] Available at: http://www.iso.org/iso/catalogue_detail.htm?Csnumber=59226 [Accessed 20 September 2016].

Jabeen, F., 2010. Eye Tracking in Web Usability: What Users Really See (Doctoral dissertation).

Jiang, Y., Olson, I.R. and Chun, M.M., 2000. Organization of visual short-term memory. Journal of Experimental Psychology: Learning, memory, and cognition, 26(3), p.683.

Jiang, Y.V., Lee, H.J., Asaad, A. and Remington, R., 2016. Similarity effects in visual working memory. Psychonomic bulletin & review, 23(2), pp.476-482.

Johnson, N.F., 1964. The functional relationship between amount learned and frequency vs. rate vs. total time of exposure of verbal materials. Journal of Verbal Learning and Verbal Behavior, 3(6), pp.502-504.

Jones M., Marsden G., Mohd-Nasir N., Boone K. And Buchanan G., 1999. Improving Web interaction on small displays, Computer Networks, Volume 31, Issues 11–16, Pages 1129-1137, ISSN 1389-1286, https://doi.org/10.1016/S1389-1286(99)00013-4.Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 683–702. doi:10.1037/0278-7393.26.3.683

Judd, T., Durand, F., and Torralba, A., 2012. A benchmark of computational models of saliency to predict human fixations.

Kahneman, D. and Egan, P., 2011. Thinking, fast and slow (Vol. 1). New York: Farrar, Straus and Giroux.

Kahneman, D., And Treisman, A., 1984. Changing views of attention and automaticity. In: R. PARASURAMAN and D. R. DAVIES (eds.), Varieties of attention. New York: Academic Press. Pp. 29–62.

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

Kärkkäinen, L. And Laarni, J., 2002, October. Designing for small display screens. In Proceedings of the second Nordic conference on Human-computer interaction (pp. 227-230). ACM.

Kim, K.J., 2017. Shape and size matter for smartwatches: effects of screen shape, screen size, and presentation mode in wearable communication. Journal of Computer-Mediated Communication, 22(3), pp.124-140.

Koehler, K., Guo, F., Zhang, S., and Eckstein, M. P., 2014. What do saliency models predict? Journal of vision, 14(3), 14-14.

Koffka, K., 1935. Principles of Gestalt Psychology. New York: Harcourt, Brace.

Köhler, W. 1929. Gestalt Psychology. New York: Liveright.

Kohli, R., 1988. Assessing interaction effects in Latin square-type designs. International Journal of Research in Marketing, 5(1), pp.25-37.

Konkle, T., Brady, T.F., Alvarez, G.A. and Oliva, A., 2010. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. Journal of Experimental Psychology: General, 139(3), p.558.

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I. and Schlesewsky, M., 2013. Subjective impressions do not mirror online reading effort: Concurrent EEG- eye-tracking evidence from the reading of books and digital media. PloS one, 8(2), p.e56178.

Kubovy, M. And Van Den Berg, M., 2008. The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. Psychological Review, 115, pp.131–154.

Lagunas, M., Garces, E., and Gutierrez, D., 2018. Learning icons appearance similarity. Multimedia Tools and Applications. 10.1007/s11042-018-6628-7.

Lam, H. and Baudisch, P., 2005. Summary thumbnails: readable overviews for small screen web browsers. In Proc. of SIGCHI, New York, NY, USA, 2005. ACM.

Lamy, D., Segal, H. And Ruderman, L., 2006. Grouping does not require attention. Perception & Psychophysics, 68, pp.17–31.

Landriscina, F., 2013. Simulation and learning. London: Springer Science & Business Media.

Leacock, C. and Chodorow, M., 1998. Combining local context and WordNet sense similarity for word sense identification. WordNet, An Electronic Lexical Database. The MIT Press.

Lewis, J.P., Rosenholtz, R., Fong, N. and Neumann, U., 2004, August. VisualIDs: automatic distinctive icons for desktop interfaces. In ACM Transactions on Graphics (TOG) (Vol. 23, No. 3, pp. 416-423). ACM.

Li, J., Qian, J. and Liang, F., 2018. Evidence for the beneficial effect of perceptual grouping on visual working memory: an empirical study on illusory contour and a meta-analytic study. Scientific reports, 8(1), p.13864.

Li, Y., McLean, D., Bandar, Z.A. and Crockett, K., 2006. Sentence similarity based on semantic nets and corpus statistics. IEEE Transactions on Knowledge & Data Engineering, (8), pp.1138-1150.

Lin, D., 1993. Principle-based parsing without overgeneration. In 31st annual meeting of the association for computational linguistics.

Lin, P.H. and Luck, S.J., 2009. The influence of similarity on visual working memory representations. Visual Cognition, 17(3), pp.356-372.

Lin, T.S. And Lai, C.S., 2013. The recognition and comprehension on application icons on mobile devices. 5th IADSR, pp.4069-4078.

Lintean, M. and Rus, V., 2012, May. Measuring semantic similarity in short texts through greedy pairing and word semantics. In Twenty-Fifth International FLAIRS Conference.

Littlejohn, S.W. 1999, Theories of Human Communication (6th ed.). Belmont, CA: Wadsworth.

Liversedge, S.P. and Blythe, H.I., 2007. Lexical and sublexical influences on eye movements during reading. Language and Linguistics Compass, 1(1-2), pp.17-31.

Loftus, G.R., 1972. Eye fixations and recognition memory for pictures. Cognitive Psychology, 3(4), pp.525-551.

Logie, R.H., 1996. The seven ages of working memory. In: J.T.E. Richardson, R.H. Logie, E.R. Stoltzfus, L. Hasher, R.T. Zacks, And R.W. Engle (eds.), Working memory and human cognition. New York: Oxford University Press. Pp. 31-59.

Long, J.D. and Cliff, N., 1997. Confidence intervals for Kendall's tau. British Journal of Mathematical and Statistical Psychology, 50(1), pp.31-41.

Mack, A. and Rock, I., 1998. Inattentional blindness. Cambridge, MA: MIT Press.

Mack, A., Tang, B., Tuma, R., Kahn, S., And Rock, I., 1992. Perceptual organization and attention. Cognitive Psychology, 24, pp.475–501.

Major, J.C., 2008. List composition effects for masked semantic primes: Evidence inconsistent with activation accounts.

Mandler, J.M., Seegmiller, D. and Day, J., 1977. On the coding of spatial information. Memory & Cognition, 5(1), pp.10-16.

Mangun, G.R. and Hillyard, S.A., 1991. Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. Journal of Experimental Psychology: Human perception and performance, 17(4), p.1057.

Mäntyjärvi, J. And Seppänen, T., 2002. Adapting applications in mobile terminals using fuzzy context information. In: International Conference on Mobile Human-Computer Interaction, pp. 95-107. Berlin: Springer Berlin Heidelberg.

Marcus, A., 1996. Icon and symbol design issues for graphical user interfaces: International users interface. London: John Wiley & Sons, Inc.

Maurer, M.E., Hausen, D., De Luca, A. and Hussmann, H., 2010, October. Mobile or desktop websites? website usage on multitouch devices. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (pp. 739-742). ACM.

McDougald, B. R., and Wogalter, M. S., 2014. Facilitating pictorial comprehension with color highlighting. Applied Ergonomics, 45(5), 1285–1290. https://doi.org/10.1016/j.apergo.2013.05.008.

McDougall, S.J. and Reppa, I., 2008, September. Why do I like it? The relationships between icon characteristics, user performance and aesthetic appeal. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 52, No. 18, pp. 1257-1261). Sage CA: Los Angeles, CA: SAGE Publications.

McDougall, S.J., Curry, M.B. and de Bruijn, O., 2001. The effects of visual information on users' mental models: An evaluation of Pathfinder analysis as a measure of icon usability. International Journal of Cognitive Ergonomics, 5(1), pp.59-84.

McDougall, S.J., de Bruijn, O. and Curry, M.B., 2000. Exploring the effects of icon characteristics on user performance: The role of icon concreteness, complexity, and distinctiveness. Journal of Experimental Psychology: Applied, 6(4), p.291.

McEvoy, P. and Richards, D., 2006. A critical realist rationale for using a combination of quantitative and qualitative methods. Journal of research in nursing, 11(1), pp.66-78.

Metzger, W., Spillmann, L.T., Lehar, S.T., Stromeyer, M.T. and Wertheimer, M.T., 2006. Laws of seeing. MIT Press.

Moore, C. M., And Egeth, H., 1997. Perception without attention: Evidence of grouping under conditions of inattention. Journal of Experimental Psychology: Human Perception & Performance, 23, pp. 339–352.

Moreland, R.L. and Zajonc, R.B., 1982. Exposure effects in person perception: Familiarity, similarity, and attraction. Journal of Experimental Social Psychology, 18(5), pp.395-415.

Moyes, J., 1995. Putting icons in context: the influence of contextual information on the usability of icons (Doctoral dissertation, University of Glasgow).

Murdock Jr, B.B., 1960. The immediate retention of unrelated words. Journal of Experimental Psychology, 60(4), p.222.

Murphy, J., Hofacker, C. And Mizerski, R., 2006. Primacy and recency effects on clicking behavior. Journal of Computer Mediated Communication, 11(2), pp.522-535.

Neisser, U., 1967. Cognitive psychology. Englewood Cliffs, NJ: Prentice-Hall.

Noizet, G. And Pynte, J., 1976. Implicit labelling and readiness for pronunciation during the perceptual process. Perception, 5(2), pp. 217-223.

Norman, D.A., 1966. Acquisition and retention in short-term memory. Journal of Experimental Psychology, 72(3), p.369.

Olsen, W., 2002, June. Dialectical triangulation and empirical research. In 6th IACR Annual Conference, University of Bradford.

Osman, A. and Moore, C.M., 1993. The locus of dual-task interference: psychological refractory effects on movement-related brain potentials. Journal of Experimental Psychology: Human Perception and Performance, 19(6), p.1292.

Oulasvirta, A., 2004. Human-computer interaction in mobile context: a cognitive resources perspective. Unpublished licentiate thesis. University of Helsinki, Finland.

Palmer, J. and Ames, C.T., 1992. Measuring the effect of multiple eye fixations on memory for visual attributes. Perception & Psychophysics, 52(3), pp.295-306.

Palmer, S.E., 1999. Vision science: Photons to phenomenology. Cambridge, MA: MIT Press.

Payne, P.R and Starren, J.B., 2005. Quantifying visual similarity in clinical iconic graphics. Journal of the American Medical Informatics Association, 12(3):338-45. doi: 10.1197/jamia.M1628

Pedersen, T., Patwardhan, S. and Michelizzi, J., 2004, May. WordNet: Similarity: measuring the relatedness of concepts. In Demonstration papers at HLT-NAACL 2004 (pp. 38-41). Association for Computational Linguistics.

Peirce, C.S., Weiss, P. And Hartshorne, C., 1931. Collected papers of Charles Sanders Peirce. London: Belknap Press of Harvard U.P.

Peterson, D.J. And Berryhill, M.E., 2013. The Gestalt principle of similarity benefits visual working memory. Psychonomic Bulletin & Review, 20(6), pp.1282-1289.

Pollack, I. and Johnson, L., 1963. Continuing memory span for digits. Perceptual and Motor Skills, 17(3), pp.731-734.

Pu, P. (n.d.). Semiotics and Icon Design. Retrieved December 12, 2018 from the EPFL website,https://ldc.usb.ve/~vtheok/cursos/ci4325/presentaciones/semiotics-icon-design.pdf

Queiroz, J. And Merrell, F., 2006. Semiosis and pragmatism. Sign Systems Studies, 34, pp.37-64.

Quinlan, P.T. And Wilton, R.N., 1998. Grouping by proximity or similarity? Competition between the Gestalt principles in vision. Perception, 27, pp.417–430.

Rada, R., Mili, H., Bicknell, E. and Blettner, M., 1989. Development and application of a metric on semantic nets. IEEE transactions on systems, man, and cybernetics, 19(1), pp.17-30.

Raptis, D., Tselios, N., Kjeldskov, J. And Skov, M.B., 2013, August. Does size matter? investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. In Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services (pp. 127-136). ACM.

Reber, R., Winkielman, P. and Schwarz, N., 1998. Effects of perceptual fluency on affective judgments. Psychological science, 9(1), pp.45-48.

Reed, S.K., 2013. Psychological processes in pattern recognition. Academic Press.

Reisner, P., 1981. Formal grammar and human factors design of an interactive graphics system. IEEE Transactions on Software Engineering, (2), pp.229-240.

Rock, I., 1986. The description and analysis of object and event perception. In: K. R. BOFF, L. KAUFMAN and J. P. THOMAS (eds.), Handbook of perception and human performance. New York: Wiley. Pp. 1–46.

Rogers, Y. and Oborne, D.J., 1987. Pictorial communication of abstract verbs in relation to human-computer interaction. British Journal of psychology, 78(1), pp.99-112.

Rogers, Y., 1986. Evaluating the meaningfulness of icon sets to represent command operations. In: Human Factors in Computing Systems: Proceedings of CHI 86, pp. 586–603.

Rohrer, C., 2014. When to use which user-experience research methods. Nielsen Norman Group.

Roibás, A. C., 2002. Computational semiotics for 3G. [online] Available at: <https://www.jiscmail.ac.uk/cgibin/webadmin?A3=ind0209&L=FWS_WP4&E=base64&P=6521720&B=_%3D_nextpart_000_01C26498.8BAAD330&T=application%2Fmsword;%20name=%22computational%20semiotics%20for%203G%20guis.doc%22&N=computational%20semiotics%20for%203G%20guis.doc&attachment=q.> [Accessed 20 August 2016].

Rubin, E., 1915. Synsoplevede figurer.

Run, S., December 04, 2014. Computing Semantic Similarity for Short Sentences.[online] Available at: <http://sujitpal.blogspot.com/2014/12/semantic-similarity-for-short-sentences.html>

Rus, V., Lintean, M., Banjade, R., Niraula, N. and Stefanescu, D., 2013. Semilar: The semantic similarity toolkit. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations (pp. 163-168).

Russell, C. And Driver, J., 2005. New indirect measures of "inattentive" visual grouping in a change-detection task. Perception & Psychophysics, 67, pp.606–623.

Salter, K.C. and Fawcett, R.F., 1993. The ART test of interaction: a robust and powerful rank test of interaction in factorial models. Communications in Statistics-Simulation and Computation, 22(1), pp.137-153.

Schade A, 2017. Big Pictures on Small Screens: Remove, Resize or Reorganize, [online] Available at: <https://www.nngroup.com/articles/big-pictures-small-screens/>[Accessed on 18th September 2018].

Schröder, S., and Ziefle, M., 2008. Effects of Icon Concreteness and Complexity on Semantic Transparency: Younger vs. Older Users. Computers Helping People with Special Needs: 11th International Conference, ICCHP 2008, Linz, Austria, July 9-11, 2008. Proceedings (pp.90-97). doi:10.1007/978-3-540-70540-6_12

Scott, D., 2007. Resolving the quantitative–qualitative dilemma: a critical realist approach. International Journal of Research & Method in Education, 30(1), pp.3-17.

Setlur, V. and Mackinlay, J.D., 2014, April. Automatic generation of semantic icon encodings for visualizations. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 541-550). ACM.

Setlur, V., Albrecht-Buehler, C., A. Gooch, A., Rossoff, S. and Gooch, B., 2005, September. Semanticons: Visual metaphors as file icons. In Computer Graphics Forum (Vol. 24, No. 3, pp. 647-656). Oxford, UK and Boston, USA: Blackwell Publishing, Inc.

Shen, Z., Xue, C. And Wang, H., 2018. Effects of Users' Familiarity with the Objects Depicted in Icons on the Cognitive Performance of Icon Identification. i-Perception, 9(3), p.2041669518780807.

Shima, H., 2013. WS4J Demo. Available at: http://ws4jdemo.appspot.com

Shrestha, S. 2007, Mobile Web Browsing: usability study Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology, ACM, Singapore.

Silfver, P., 2012. The elements of navigation. [online] Available at: <https://www.smashingmagazine.com/2012/03/the-elements-of-navigation/> [Accessed 12 December 2016].

Simpson, T. and Dao, T., 2005. WordNet-based semantic similarity measurement. The Code Project. com.

Singer, M.H. and Lappin, J.S., 1976. Similarity: Its definition and effect on the visual analysis of complex displays. Perception & Psychophysics, 19(5), pp.405-411.

Slimani, T., 2013. Description and evaluation of semantic similarity measures approaches. arXiv preprint arXiv:1310.8059.

Slimani, T., Yaghlane, B.B. and Mellouli, K., 2006. A new similarity measure based on edge counting. Proceedings of the World Academy of Science, engineering and Technology, 17, p.3.

Soegaard, M., 2010. Gestalt principles of form perception. Interaction-Design, p.8.

Speroni, M., 2006. Mastering the semiotics of information-intensive web interfaces. Thesis. Universitأ della Svizzera italiana.

Stammers, R.B. and Hoffman, J., 1991, September. Transfer between icon sets and ratings of icon concreteness and appropriateness. In Proceedings of the Human Factors Society Annual Meeting (Vol. 35, No. 5, pp. 354-358). Sage CA: Los Angeles, CA: SAGE Publications.

Statista, 2018. Mobile internet traffic as percentage of total web traffic in April 2018, by region [online] Available at :< https://www.statista.com/statistics/306528/share-of-mobile-internet-traffic-in-global-regions/>, [Accessed on 15th September 2018].

Stevens, J.E., Robinson, A.C., and MacEachren, A.M. 2013, “Designing map symbols for mobile devices: Challenges, best practices, and the utilization of skeuomorphism”, 26th International Cartographic Conference, Pennsylvania State

University. Available at https://icaci.org/files/documents/ICC_proceedings/ICC2013/_extendedAbstract/356_proceeding.pdf

Stickel, C., Pohl, H.-M. And Milde, J.-T., 2014. Cutting edge design or a beginner's mistake?- semiotic inspection of ios7 icon design changes. In: International Conference of Design, User Experience, and Usability. Springer International Publishing. Pp. 358-369.

Stotts, D.B., 1998, October. The usefulness of icons on the computer interface: Effect of graphical abstraction and functional representation on experienced and novice users. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 42, No. 5, pp. 453-457). Sage CA: Los Angeles, CA: SAGE Publications.

Sunday, M., 2014. Determining a Correlation Between Individual Differences in Eye Movements and Working Memory.

Taylor, J., 1995, Linguistic Categorization: Prototypes in Linguistic Theory (2nd Ed.). Oxford: Oxford University Press.

Tepfenhart, W.M. and Sowa, J.F., 1994. Conceptual Structures: Current Practices: Second International Conference on Conceptual Structures, ICCS'94, College Park, Maryland, USA, August 16-20, 1994. Proceedings (Vol. 835). Springer Science & Business Media.

Tijus, C., Barcenilla, J., De Lavalette, B. C. And Meunier, J.-G., 2007. The design, understanding and usage of pictograms. Studies in Writing, 21, p.17.

Tobii Technology, 2012, How do Tobii Eye Trackers work? [online] Available at: <http://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/>[Accessed on 10th April 2018].

Tobii Technology, 2019, Tobii IS5 Platform, [online] Available at <https://www.tobii.com/tech/products/platforms/>

Tocher, K.D., 1963. The Art of Simulation', English Universities Press.

Trapp, A.K. and Wienrich, C., 2018. App icon similarity and its impact on visual search efficiency on mobile touch devices. Cognitive Research: Principles and Implications. 3:39. https://doi.org/10.1186/s41235-018-0133-4

Treisman, A.M. and Gelade, G., 1980. A feature-integration theory of attention. Cognitive psychology, 12(1), pp.97-136.

Trogu, P., 2015. Working memory and background knowledge: Cognitive science in the design classroom. Formakademisk, 8(1), pp.1-17.

Tversky, B., 1973. Encoding processes in recognition and recall. Cognitive psychology, 5(3), pp.275-287.

Tversky, B., 1974. Eye fixations in prediction of recognition and recall. Memory & Cognition, 2(2), pp.275-278.

Vogel, E.K., Woodman, G.F. and Luck, S.J., 2005. Pushing around the locus of selection: Evidence for the flexible-selection hypothesis. Journal of cognitive neuroscience, 17(12), pp.1907-1922.

Wang, D.L., 2003. Visual scene segmentation. In: ARBIB, M.A., (ed.), The handbook of brain theory and neural networks. 2nd ed., Cambridge, MA: MIT Press, pp. 1215-1219.

Wedel, M. and Pieters, R., 2000. Eye fixations on advertisements and memory for brands: A model and findings. Marketing science, 19(4), pp.297-312.

Wertheimer, M., 1923. Laws of organization in perceptual forms. A source book of Gestalt Psychology.

Wiedenbeck, S., 1999. The use of icons and labels in an end user application program: an empirical study of learning and retention. Behaviour & Information Technology, 18(2), pp.68-82.

Wiswede, D., Rüsseler, J. and Münte, T.F., 2007. Serial position effects in free memory recall—An ERP-study. Biological psychology, 75(2), pp.185-193.

Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J., 2011, May. The aligned rank transform for nonparametric factorial analyses using only anova procedures.

In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 143-146). ACM. ARTool available online at: http://depts.washington.edu/acelab/proj/art/index.html.

Wolff, J. S. And Wogalter, M. S., 1998. Comprehension of pictorial symbols: Effects of context and test method. Human Factors: The Journal of the Human Factors and Ergonomics Society, 40, pp.173-186.

Woodman, G. F., Vecera, S. P., and Luck, S. J., 2003. Perceptual organization influences visual working memory. Psychonomic Bulletin & Review, 10, 80–87. doi:10.3758/BF03196470.

Xu, Y., and Chun, M. M., 2007. Visual grouping in human parietal cortex. Proceedings of the National Academy of Sciences, 104, 18766–18771. doi:10.1073/pnas.0705618104.

Xu, Y., 2002. Encoding color and shape from different parts of an object in visual short-term memory. Perception & Psychophysics, 64, pp.1260–1280.

Xu, Y., 2006. Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. Perception & Psychophysics, 5, pp.815–828.

Yarbus, A.L., 1967. Eye movements during perception of complex objects. In Eye movements and vision (pp. 171-211). Springer, Boston, MA.

Zakia, R.D. And Nadin, M., 1987. Semiotics, advertising and marketing. Journal of Consumer Marketing, 4(2), pp.5-12.

Zhang, J. and Ghorbani, A.A., 2004, October. Familiarity and Trust: Measuring Familiarity with a Web Site. In PST (pp. 23-28).

Zimmerman, M.E., 2011. Speed–accuracy tradeoff. Encyclopedia of clinical neuropsychology, pp.2344-2344.