Gauging User Tolerance to Interface Modifications Using Gestalt Principles

By

Rachelle Barrette

A thesis submitted in partial fulfillment

of the requirements for the degree of

Master of Science (MSc) in Computational Science

The Faculty of Graduate Studies

Laurentian University

Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	Gauging User Tolerance to Interface Modifications Using Gestalt Principles	
Name of Candidate Nom du candidat	Barrette, Rachelle	
Degree Diplôme	Master of Science	
Department/Program Département/Programme	Computational Sciences	Date of Defence Date de la soutenance July 07, 2021

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Ratvinder Grewal
(Supervisor/Directeur(trice) de thèse)

Dr. David Vares
(Committee member/Membre du comité)

Dr. Kalpdrum Passi
(Committee member/Membre du comité)

Dr. Savita Gandhi
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Tammy Eger, PhD
Vice-President Research (Office of Graduate Studies)
Vice-rectrice à la recherche (Bureau des études supérieures)
Laurentian University / Université Laurentienne

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Abstract

Interface updates are more than all too common in the twenty-first century. Constantly, users are bombarded by new updates incurring changes, and new functionality to not one but many of their favourite applications and devices. How changes affect the users can be studied in various settings for numerous reasons. This paper examines a situation where the health and safety of individuals may be at risk; specifically, the patients in the care of nurses and when the input of the medical equipment that the nurses' use is changed. It is theorized to have unforeseen consequences on the nurse, leading to critical, even fatal errors to the patients in their care. This research aims to outline what interface changes can be made and how these changes affect the user in terms of performance and cognitive load. To accomplish this, an experiment is conducted where participants' play a simple memory game, and the inputs are changed during gameplay. Simultaneously, performance scores, time, number of errors, and cognitive load are tracked to infer to what degree users are affected by the change. It is hypothesized that the changes tested are not significant enough to elicit a large reaction from the user, and this was found to be true in the performance and cognitive load scores collected, other than a few exceptions. Given this, we can state that user interface modifications based on the Gestalt Principles of Figure-Ground, Proximity, and Closure did not, in this case, elicit a significant effect from the user. Therefore, changes like those tested could be implemented on an interface in a similar scenario and given to users with little concern for negative repercussions. Although, to generalize to a larger context, the experiment would need to be reconducted with more participants to elicit a significant effect from the data.

Acknowledgements

Firstly, I would like to thank Thomas Malley for his continuous support and open ears over the last few years; your help and support have been paramount to completing this document.

Secondly, I would like to thank my supervisor, Dr. Ratvinder Grewal, for his support and encouragement throughout my research. Thank you for the time and effort you take to help your students grow and prosper.

Thirdly, thank you to all of the CHILL group members, current and past, for the weekly discussions, your suggestions and various perspectives that have helped me throughout this journey.

Lastly, but definitely not least, thank you to my sisters. Your constant support and encouragement have been so very appreciated. Thank you for always being there for me when I have needed you both the most.

Finally, to anyone I may be forgetting that have helped me along the way, thank you.

Funding Acknowledgements

This project is funded in part by a Seed Grant provided by the Centre for Research in Occupational Safety and Health (CROSH).

centre for research in occupational safety and health at Laurentian University

CROSH CRSST

centre de recherche sur la santé et sécurité au travail à l'Université Laurentienne

"If I have seen further, it is by standing upon the shoulders of giants."

- Sir Isaac Newton

Table of Contents

List of Figures

List of Tables

List of Acronyms

CL – Cognitive Load

ECG or EKG - Electrocardiogram

EEG - Electroencephalogram

EMG - Electromyography

HCI – Human-Computer Interaction

NASA-TLX - National Aeronautics and Space Administration–Task Load Index

OW - Overall Workload Scale

RQ – Research Question

SWAT - Subjective Workload Assessment Technique

UI – User Interface

UX – User Experience

Chapter 1 - Introduction

Interface updates are more than all too common in the twenty-first century. Constantly, users are bombarded by new updates, changes, and new functionality, to not one but many of their favourite applications and devices. This is shown in an article by McIlroy [37]: "We studied the frequency of updates of 10,713 mobile apps (the top free 400 apps at the start of 2014 in each of the 30 categories in the Google Play store). We find that a small subset of these apps (98 apps representing ˜1 % of the studied apps) are updated at a very frequent rate — more than one update per week and 14 % of the studied apps are updated on a bi-weekly basis (or more frequently)." Therefore, each and every week users are required to adapt to the new updates.

There has been little consideration of how the user reacts to these changes in terms of performance and cognitive load. The following authors' have been surveyed [33], [34], [35], [30], [6], [39], [43] and [17], as they have tested how various changes affect the user, including keyboard layouts, visual search elements and gestalt principles. From this research, it has been found, in general, that following the users' preference [33] and highlighting the required buttons [43] reduces the disruptions caused to users' and increases their performance.

The importance of this research lies in many situations, but the one considered here is the study of health and safety in the workplace, specifically with nurses, and in the equipment that they use daily. For example, if we consider an infusion pump, a nurse may be used to working with a specific pump from a specific manufacturer. However, there may be instances where they encounter a different pump in their workplace. What can we expect to happen when they use this new, different pump? Moreover, we also must consider the environmental effects surrounding a nurse. When considering a working nurse, who is overloaded on a 12-hour workday, they are juggling many patients throughout this work shift, burdened by the many visual and audio cues around them by the telecom, pieces of equipment and other nurses and doctors needing them or their input; how do these environmental effects factor into their mental state and performance on the job?

My research aims to outline what interface changes can be made and how these changes affect the user in terms of performance and cognitive load. To accomplish this, an experiment is conducted where participants' play a simple memory game, and the inputs are changed during gameplay. Simultaneously, performance scores, time and number of errors, and cognitive load are tracked to infer to what degree they are affected by the change.

In my thesis, relevant literature is reviewed. To begin, we first explore cognitive load theory and how cognitive load affects the user during the conduction of a task—followed by the exploration of the transition of novice to expert users. Lastly, my literature review covers current studies that have tested the effects of various user interface modifications. Subsequently, the experimental design is explained. Finalizing with the results, leading to a discussion and conclusion of the results.

Chapter 2 – Literature Review

In this chapter, various academic articles are reviewed. Firstly, Mental Models and User Experience are analyzed to understand its' relevance during the use of software. Secondly, Cognitive Load is reviewed to understand what role it plays within users and their daily technological activities. Lastly, various user interface modifications that have been tested in recent literature are examined.

2.1 Mental Models

A mental model, the conceptual model built inside a user's mind, describes the user's knowledge of a program's functionality [40]. As Don Norman stated in his book, "Different people may hold different mental models of the same item" [40]. For instance, when sampling a population of novice or casual computer users, their mental model may be as simple as "if I press the power button, the laptop will turn on, I can put in my password to unlock my laptop and begin using it." In comparison to a computer scientists' mental model will consist of many high-level concepts, including threads, processes, data moving between registers, and substantially more.

A mental model can either be weak ("shallow") or strong ("deep") [44]. A strong mental model is one that closely resembles the mental model of the designer or developer that created the object/program. A weak mental model does not hold this same link between the user and the designer's mental model. If we revisit the above example, the educated population (computer scientist) would have a “strong” or “deep” mental model, where the casual population would have a “weak” mental model.

Mental models are not static, but rather dynamic in that users can strengthen their mental models using an array of information at their disposal [40]. This can be information from sources such as instruction manuals, product reviews, pictures, education, or even conversations with friends, family or salespersons. Most importantly, the most reliable foundation of a mental model is from the use of similar objects [40]. If we reconsider the educated population, an educated individual could apply previous knowledge of a different but similar object to begin to understand the inner workings of a desktop computer or cellular device. Even though these devices will differ in size and complexity, the essential concepts of how they work do not vary a great deal. This example creates the basis for the idea of knowledge transfer [51]. Even though different software’s have different features, there are still some similarities or industry standards that are shared by most software programs, such as cropping, and quick actions, for example, copy (ctrl+c), paste(ctrl+v), save(ctrl+s). This type of knowledge can be ‘transferred’ when using a new program.

2.2 User Experience

User Experience (UX) denotes the interaction surrounding the user, the interface and the context [49]. An interface succeeds when the user experience is enjoyable; to guarantee a good user experience is to design the interaction and context surrounding that interface. A patent (ISO 9241-210) [15] is in place to help designers accomplish this. This patent also provides advice on design principles to ensure a good UX. UX considers the experience with the intended use of a system and past experiences with the system. Therefore, UX englobes all interactions with any device or system [49].

The mental model of a user can affect a user's experience. The pre-cursor knowledge that a user carries can set expectations in their minds about what the user believes will happen or what the system can accomplish, which will guide their experience with a system. Because of the user's pre-cursor knowledge, it is crucial to consider a user's mental model throughout the user experience. Over time, a user's mental model can be updated based on new knowledge or experiences. UX is best modelled with Figure 1 from [49]. Prior to the user having used the system, the user can imagine the experience that they will have with the anticipated UX. At this step, their current mental model is helping the user imagine the experience that they are about to have [49]. Once they begin using the system, new information that the user is experiencing will be added to their mental model from the momentary UX to help the user understand how to use the system [49]. After the use of the system, the user will reflect on what they have experienced, update their mental model with any new information they have gained, and the cycle repeats. Lastly, over time, the user will recollect their experience within their UX. Furtherly, this cycle continuously adds more new information to their mental model [49]. This cycle is endless, as it may happen fully or partially for every experience with the same or different systems.

The context of the use of the system is also essential to consider. A user's context can change dramatically and quickly; this can include social context, whether the user is working alone or in a group, the physical context (the type of device), geographical context (where it is being used: at home, at school, at work or on the go), and finally, the intended purpose of the usage of the system. These are all factors that can significantly affect the user experience and should be considered during the design process to ensure a good user experience. [49]

Although, what happens if a user encounters a new, similar, but differently unique software program. For example, if a user were to encounter Corel Draw after using Adobe Photoshop. Given the understanding of knowledge transfer, we would assume that the user would take their stored knowledge and transfer it to the task at hand which is the new interface they are encountering. For some users, this will be a difficult, time-consuming, stressful task; knowledge transfer is not a natural skill for everyone; this is where cognitive load illuminates the situation and to better understands how much mental activity the user expends. With the understanding that every user is different, we should consider how a user's cognitive load is affected by a change.

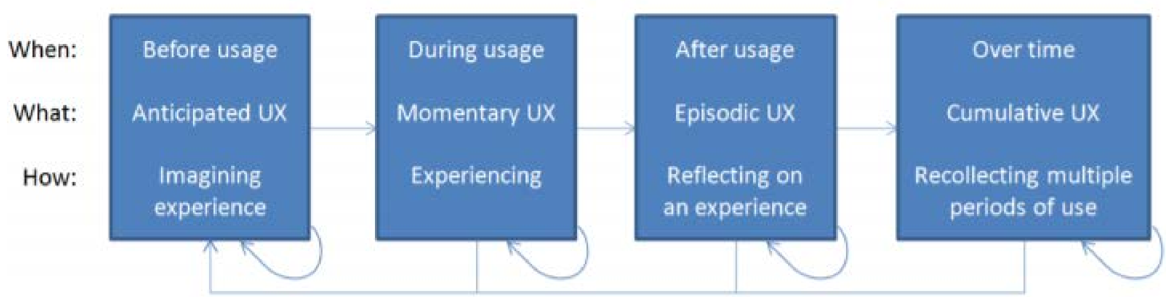


Figure 1 [49] – User Experience Cycle

Given this, it is now understood that users build mental models of the interface and therefore, software programs that they are exposed to. The countless times they see and use them affect their user experience. Therefore, it is imperative to consider how the interface affects the user experience and how it will be perceived to ensure the creation of a mental model. This is important because the knowledge the user gains will help them transfer their knowledge in the future. For all those reasons, if a mental model is built, then knowledge can be transferred. With this, we should be able to test how changes affect the users' mental model, their user experience, but first we will explore cognitive load in greater depth to better understand at what point a mental model is built.

2.3 Cognitive Load Theory

Cognitive load is an alternate term to "mental effort" or "mental workload." Cognitive load describes the amount of mental power that it takes to conduct a task. A task includes a deduction task, a problem-solving task, mathematical task, or any task that consumes some mental capacity. Cognitive load is not binary; it is not a matter of having or not having cognitive load; it is measurable from high to low or 1 to 10 if measured on a subjective scale. Cognitive load varies per person, per task, depending on the person's background, experience, and skills.

2.3.1 Cognitive Load in HCI

The following scenario is a problematic one but a common scenario for many users in the twenty-first century: when a user encounters a new technology, the user's knowledge may not always transfer from a previous technology that they have used. For example, as a UI/UX designer, we should consider the different factors that affect a user. This should be done to optimize the interface to support their understanding from when they begin to use the device and nurture them to a higher level of understanding. Cellphone carriers, such as Rogers [48] or Bell [4], carry devices mostly from three big competitor devices: Apple [2], Google [22] and Samsung [50]. These three competitors' user interfaces all look slightly different from one another but accomplish the same tasks: sending texts, making calls, installing apps. Yet, for the average user, the switch between one device to another can be very difficult. Given these differences between devices, it is crucial not only to help a user advance their mental model on one single device but also to enable them to change between such devices with the knowledge they have previously acquired. This should be done so the user does not just memorize the steps they are taking to accomplish the task but understand the steps as well. The point is that accomplishing this difficult task assures that the user experience is going well. After all, if the user can transfer their skills and understanding from one device to the next, it guarantees that the user understands how to function said device because their mental model has been formed and stored in long-term memory [42].

When the user experience is not maintained, the user's cognitive load is continuously on "high," as they are continuously struggling to understand the task they are trying to accomplish and therefore, there is no opportunity to form a mental model [42]. This situation will leave the user frustrated and uneasy as they will not know what steps to take next, which will inadvertently feed into the uneasy cycle of a poor user experience. Even if a previous mental model is formed and stored, technology is constantly changing, a user will always be required to adapt and learn new features on the fly.

Additionally, according to Nielsen's 10th heuristic, "help and documentation" [44], a user interface should be self-explanatory; it should not require additional instructions or information to understand its use and available functionality. However, this is a challenge in many regions of user interface design. In website design, this challenge is not as difficult to circumvent. A website is relatively straightforward; as stated by Garett [21], a user will typically visit a website to find information about a business, or person, service or product. Websites are used to communicate this information with the general public [21]. The underlying task is not complicated and is common across many websites since all websites serve this very similar purpose. However, in a user interface for accountancy software or medical equipment, where many functions are at play, such as infusion rate, infusion liquid, etc.., it is not easy to ensure that the user always interprets the interface as self-explainable. Additionally, as previously mentioned, the user's background knowledge, experience, confidence, anxiety level, and preferences can affect how they interpret an interface and whether the functionality needed will be apparent.

2.3.2 Cognitive Load in Health Care

Cognitive load in Healthcare poses more significant challenges than simply confusing their users and diminishing the user experience. There are even more factors that come into play when considering cognitive load in the healthcare setting. For example, if we consider nurses in an Emergency Room environment, they may have to juggle multiple patients, per patient multiple pieces of medical equipment, some of which will make noises, as well as visual cues from said equipment. Nurses also have plenty responsibilities during their work shift, such as: dosages to administer to patients, when to administer such dosages, risks to patients, the list continues. They must be prepared to handle any emergencies they may encounter and be prepared for fast problem-solving at a moment's notice. Moreover, there is more at risk in these scenarios than a bad user experience if the nurse is confused about the device they are using, this is covered in the following section.

2.3.3 The Risks of Cognitive Load

The risks of a high level of cognitive load are endless; not only does it include stress, frustration, confusion, confidence, motivation, and a weak mental model as briefly mentioned; there are also crucial risks such as errors. In the context of healthcare, or industrial workplaces, errors could cost lives.

A famous example is the airbus accident described in Don Norman's book [40]: In this accident, the flight control equipment had two modes: controlling vertical speed and angle of descent. In one case, the pilots thought they controlled the angle of descent but were controlling speed. An inappropriate entry of -3.3 for speed caused a descent too quickly and resulted in a fatal accident of both pilots. This incident resulted in a change of the input instrument to ensure that speed would always be a four-digit number and angle a two-digit number to allow both variables to be distinguishable by its input. [40]

A second example is a hospital entry system in an example, given by Koppel [31]. In this example, a hospital pharmacy may only stock ten milligram pills for a medication prescribed in twenty or thirty milligram doses to prevent overstocking an under-prescribed medication. It was found that when this prescription is entered into the system, it was not always changed from the default value populated [31]. Typically, this is because of the time pressure of the job and not enough attention is paid to the task at hand. As mentioned previously, this small error could cause severe adverse effects if the patient gets the wrong dosage. Additionally, this system in question has up to twenty screens to review all the patient's information, medications, history and more. It was found in this study that seventy-two percent of the staff reported difficulties in reading, digesting, and remembering this information. Who could remember all the information in twenty pages? [31] Miller's theory tells us that we can only remember seven plus or minus two pieces of information at one time [20]. Therefore, information spanning a multitude of pages clearly will not be absorbed.

A third example is the interface in the BMW 745i vehicle [28]. The BMW 745i's interface, for one, uses obscure abbreviations. These abbreviations impose additional cognitive load because the driver must process and understand these abbreviations while conducting the primary task of driving. Especially in this case, because the drivers are not mechanics and therefore, do not hold the same mental models, they will not understand them. Also, this vehicle comes with two keys where each key that is intended to identify two separate drivers; when that driver and their key enter the vehicle, it automatically adjusts the chair, mirror settings, and other relevant components to that driver. Regardless of the novelty of this feature, there is no simple way of distinguishing keys without modifications performed from the user. Therefore, the drivers could easily misplace and mix up these keys. Lastly, when the driver uses the GPS, the driver must move the pointer to the command to initiate route guidance. It can be safely assumed once the address is inputted into the GPS, the user intends to initiate route guidance to that location. This redundant step of initiating guidance after inputting the address is confusing, redundant, and unnecessary. [28]

There are not many examples such as these in a nursing environment which demonstrate the same story. Although, it is demonstrated by authors such as [52] that the most common error with medical devices is "use errors", where the fault is set on the nurse. Although, the process to document these is not detailed. Therefore, Silva [52] suggests a new classification method of these errors to better understand where these errors are stemming from. If these are design related, then the issue of how the designs can be changed to mitigate these errors can be addressed [52].

2.3.4 What can increase cognitive load?

The following list contains factors that can play into increasing cognitive load, since the user then must focus more on the given factor rather than the task at hand.

- Unclear Instructions
- Complex Task
- Not enough experience/practice
- Overwhelming Instructions/Interface/Feedback

- Insufficient Feedback
- Disruptions

An example of how disruptions can affect cognitive load can be found in [36]. The goal in the following article: [36] by Mark was to test the cost of disruptions, the amount of time it would take to reorient back to the original task, whether this is for same-context disruptions or out-of-context disruptions. The key results from this study are that "when out-of-context disruptions were presented, cognitive load was most prevalent, and when no disruptions were presented, the participants took their time and felt less time pressure".

2.3.5 Types of Cognitive Load

There exist three types of cognitive load: intrinsic, germane, and extraneous cognitive load. [42]

Intrinsic cognitive load is the load imposed by the task being learned and its' complexity; the more complex the task, the additional load it will impose. When speaking of complexity, the theory of interacting elements comes into play here, the more interacting elements, the more complex the task. [42]

Extraneous cognitive load is the load imposed given the design of the material trying to be learned. Therefore, the harder the task is to understand, the additional cognitive load it imposes. We can control this by revising and adapting how the interface presents the information to the user by optimizing the user experience and improving its usability. [42]

Lastly, like extraneous, Germane cognitive load is the load imposed by the design of the material. The difference between extraneous and germane is that germane also handles constructing and updating schemas. A schema, also known as a mental model, is the model that contains an individual's understanding of how a device or software functions, how to accomplish a task [40]. Therefore, Germane cognitive load is the part that handles the creation of the schema and storing it into long-term memory for later use. [42]

These types of cognitive load also interfere with each other. For one, these three loads are mutually entangled; the total of the three cannot exceed the total working memory available [42]. Secondly, Intrinsic, and extraneous are entangled between each other; if one is high and the other is not, no overload should occur. If both are high, an overload will occur [42]. This can be seen in Figure 2, a) and b) [56]. Moreover, when the interface ensures that all three operate optimally, the three work in harmony, optimizing the working memory resources and optimizing germane cognitive load to create and store schemas. This can be seen in Figure 2, c) [56]. Lastly, resources are allocated to intrinsic first, then extraneous, followed by germane [42]. Meaning that to begin, the task is processed, then the design of the task and followed by the schema. Therefore, if an overload occurs, a schema will never be made, and therefore, the experience will be jeopardized. Cognitive load must be reduced before a schema can be formed [42].

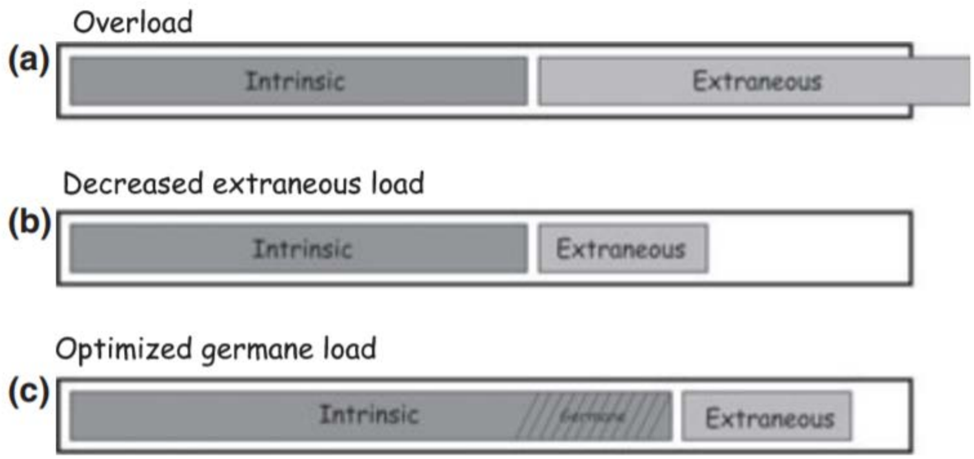


Figure 2 [56] – Cognitive Load Interconnectedness

2.3.6 Theories in Cognitive Load

Many theories exist in the realm of cognitive load to help explain how cognitive load is induced on a learner. The following explains a few of the most important ones.

2.3.6.1 Element Interactivity

The theory of interacting elements can be described in levels from low to high [8]. Low-element interactivity would exist when a task is presented with few inter-tangled components to complete or understand [8]. High-element interactivity exists when many components need to be understood in combination with each other to complete the task at hand [8].

2.3.6.2 Expertise Reversal Effect

A complementary theory to the theory of interacting elements is the expertise reversal effect. The expertise reversal effect states that a user's expertise level needs to be considered when a task is designed to facilitate learning [8]. For a beginner learner, a walkthrough approach may facilitate learning most easily. Although, for an expert learner, with plenty of knowledge of the task at hand, a free to roam platform may be best to allow them to utilize their prior knowledge to complete the task at hand; where a walkthrough approach may prohibit their progress, slowing them down and frustrating them. [7]

An example of these two theories was tested in [7], where an experiment was conducted to confirm the expertise reversal effect. The authors hypothesized that students would perform better if given the proper level of instructions; for example, if a beginner is given an isolated elements task and an expert is given an interactive elements task [7]. Two conditions were tested: an interacting elements group where the excel sheet presented consisted of the numerical data needed for the calculation and one final answer cell, and a second, isolated elements group had multiple cells to fill in that were steps to the final answer. This experiment concluded that the novice learners scored lower when presented with the interacting elements task than the isolated elements and the experts scored lower on the isolated elements task than the interacting task. The reverse could also be seen where novice scored highest on the isolated elements task and the experts scored highest on the interacting elements task. Therefore, this shows that when the learner is assigned to task instruction class that best fits their level of knowledge, novice → isolated elements and experts → interacting elements, they perform best. [7]

2.3.7 How to collect cognitive load data?

To collect valid cognitive load data, one must use a or multiple of the methods described below to ensure sound data collection. There are four methods as stated by Brunken [9]: direct, indirect, objective, and subjective.

Indirect evaluates cognitive load by looking at time, performance, and other measures such as these. This method is questionable because we cannot assure that a fluctuation in these measures is directly related to an increase or decrease in cognitive load. An example of this can be seen in the study by Mark [36] as explained above. This study investigated the cost of disruption in a simulation-type experiment [36]. In this study, time, and number of errors to draft the emails were collected to infer results indirectly.

Direct is through the n-task paradigm described in [9]. This method entails using a secondary task to evaluate cognitive load. This method's idea is to present a secondary task to the user, which would force the user to split their attention between both tasks. Two implementations of this paradigm exist. Firstly, it can be used to purposely create cognitive load to measure the performance of the primary task. In this case, the performance of the primary task should be lower than when the primary task is evaluated alone. Secondly, it can be used to measure the load imposed by the first task. If the cognitive load of the first task increases or is high, this will be seen in the performance of the secondary task. [9]

This method presents some challenges. Firstly, the secondary task must "use the same cognitive resources as the primary task," stated [9]. For example, if the first task is a monitoring task, the secondary must be one as well. Secondly, "the secondary task's performance measure must be reliable and valid" [9]. For example, the performance could be evaluated as reaction time. Lastly, the secondary task must be easy enough not to impede the user's learning processes of the primary task at hand.

In previous research by Brunken [9], the authors have conducted two experiments to evaluate the dual-task approach for measuring cognitive load. As Brunken [9] stated, "the results of these studies clearly showed the feasibility of our dual-task approach."

Haji [24] also conducted an experiment where a dual-task measure was used to measure cognitive load during a knot-tying task with medical students. The results showed that as novice knot-tiers developed a mental model (schema) of the required task, their performance in this task began to increase, and intrinsic load began to decrease, which is also in line with cognitive load theory's assumption. The dual-task reaction time results and the subjective rating scales collected confirm this theory. This theory confirmation means that their findings support the idea that subjective ratings and reaction time-based secondary (dual) task measures of cognitive load can be used to analyze intrinsic cognitive load changes. [24]

The **objective** method uses physiological measures such as an electroencephalogram (EEG), electrocardiogram (ECG or EKG) or Electromyography (EMG) to track cognitive load. An example of a study utilizing an objective method is a study conducted by Chen [13]. In this study, the authors modified a cellular device to recognize various user states using a combination of heart rate and movement tracking. These states are recognized by the data collected through the EEG and the ECG. Once these states were recognized, the device would change its' state to match the user's state. A six-person study was conducted; from this, it was found that the appropriate state was assigned 83% of the time [13]. Although a six-person study is not extensive enough to prove this method's feasibility, it is at least a beginning to show the possibility of combining physiological measures to assess a user's cognitive load state.

The **subjective** method uses rating scales to evaluate cognitive load. This method is also questionable because it requires the user to evaluate themselves and choose their cognitive load level. Therefore, one user's perception of a level 3 cognitive load may differ from the next, implicitly skewing the results.

There exist many scales to rate cognitive load, but these are the five most common: Modified Cooper Harper Scale [26], NASA-Task Load Index [26], Overall Workload Scale [26], Subjective Workload Assessment Technique [26] and Paas's Scale [41].

1) Modified Cooper-Harper Scale (MCH)

The Modified Cooper-Harper Scale is a 10-point rating scale that has the intention for the participant to provide an overall rating of workload. A decision tree is provided for the "rater" to use as a guideline of workload level to determine where they would situate themselves on the scale of the task in question. [26] An example is given in Figure 3.

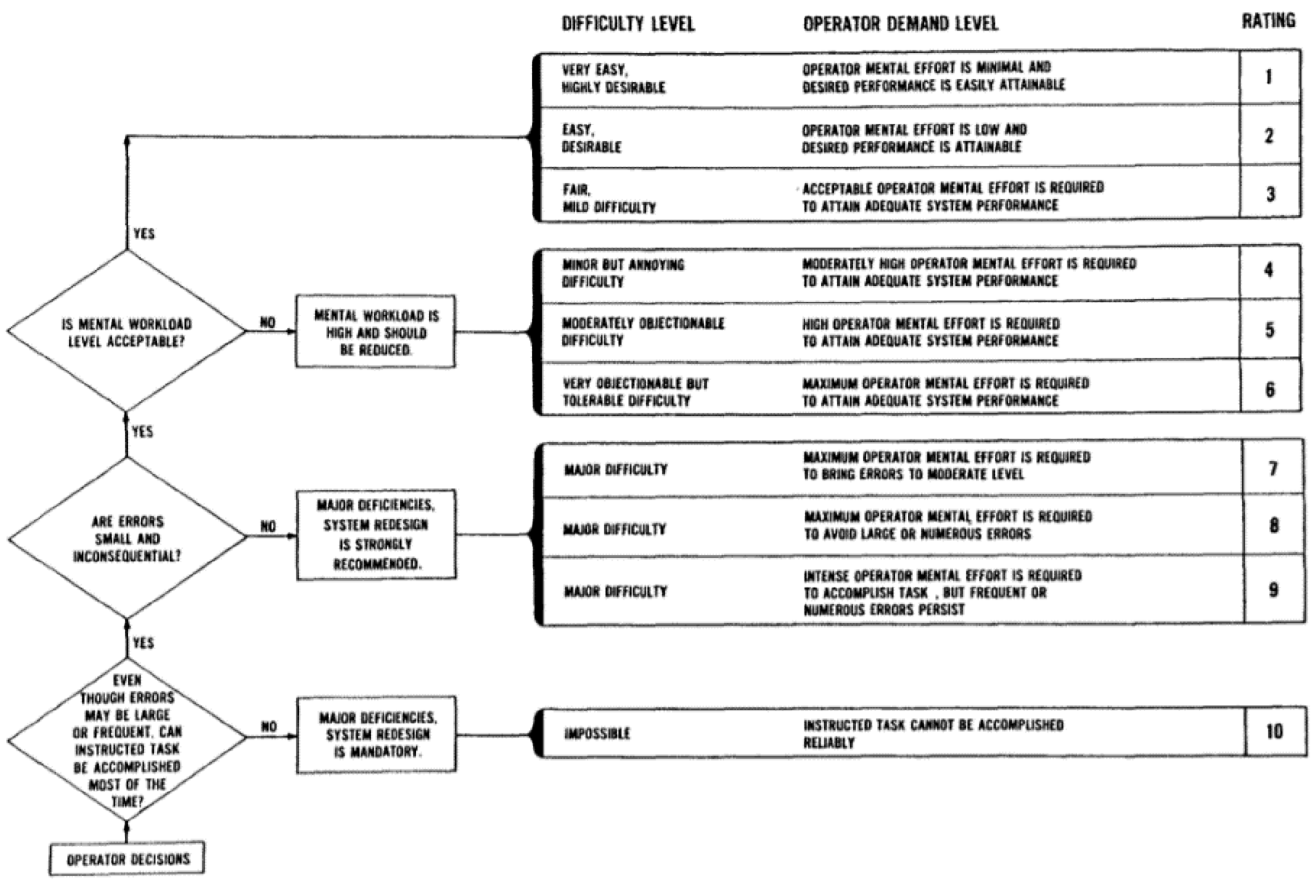


Figure 3 [57] – Decision Tree Used in the Modified Cooper Harper Scale

2) <u>NASA-Task Load Index (NASA-TLX)</u>

This index uses six dimensions: "mental demand, physical demand, temporal demand, performance, effort and frustration." The NASA-Task Load Index consists of two components: the first is a "twenty-step bipolar scale (two ended) from 0 to 100" for each dimension's ratings, and the second is comparisons between every possible combination of two of the six dimensions. The participant is shown each possible combination of two dimensions, and they are asked to choose which dimension was most prevalent to them throughout the experiment. A count is then conducted of each time each dimension was chosen. Lastly, an adjusted rating score is calculated by multiplying the workload score for each dimension from the scales by the paired comparison weights. This adjusted rating calculation is repeated for each task [25] [26].

3) <u>Overall Workload Scale (OW)</u>

The Overall Workload scale is a subjective rating scale on a 100-point scale, giving an overall workload rating. A "0 on this scale represents a very low workload, and 100 represents a very high workload". [26]

4) <u>Subjective Workload Assessment Technique (SWAT)</u>

The Subjective Workload Assessment Technique is a subjective rating scale that utilizes three dimensions: time load, mental effort load and psychological stress load, and three levels: low, medium, and high, which gives a rating for the workload [26]. There are three steps in SWAT:

1. Scale Development "All combinations of the three levels (low, medium and high) are contained on 27 cards. Each operator sorts the cards into the rank order that reflects his or her perception of increasing workload." [26]

2. Event-Scoring In this step, each task is rated for each dimension (time load, mental effort load and psychological stress load). [26]

3. Lastly, similarly to the NASA-TLX survey, each rating for each dimension is transformed into an overall score using the scale developed in 1. [26]

5) Paas's Scale

In this subjective rating scale, participants rate the perceived amount of mental effort given to a task on a "9-point symmetrical category scale" where "1" corresponds to "very, very low mental effort" and a "9" corresponds to "very, very high mental effort." [41]

2.4 Novice to Expert

To manage a person's cognitive load during the conduction of a task, we must understand at what stage of proficiency this person is. If they are new to the task, they will need more help, such as isolated tasks, as stated in the study by Blayney [7]. Dreyfus [16] suggested a five-stage process that describes the steps experienced in learning and acquiring a new skill. A process such as the one Dreyfus [16] suggested can help keep track of the person's level and adapt the interface and the task accordingly. The five-stage process that Dreyfus [16] suggested is as follows:

1) Novice

   Dreyfus [16] suggests at this stage that the novice can understand the task if it is presented in an isolated fashion as “context-free features.” The novice can also function if given steps to follow concerning these “features”. Lastly, to improve, the novice’s actions must be observed and corrected to teach the novice where to change their actions to complete the task correctly. [16]

2) Competent

   To become competent is to be in a situation where the novice has been able to cope, with instructor or supervisory help, with various situations arising from the task at hand. These coped with situations are referred to as “aspects”. Principles or guidelines are created, by the instructor or supervisor, based on the aspects encountered to give the competent a set of actions to follow when encountering these aspects. [16]

3) Proficient

With increased practice, the proficient will handle an increasing number of aspects while learning more and more each time. As the proficient continues to practice, schemas begin to form to allow them to recognize similar events in the future and remember what actions are needed, what principle is needed, and how to cope with these as they are encountered. [16]

4) Expert

   The expert has experienced many aspects at this stage and has learned the corresponding principles to cope with these. At this point, the expert has built an extensive reservoir of schemas, which has allowed the expert to build a sense of intuitiveness to situations; their intuition allows them to determine an appropriate action to a situation when they arise. [16]

5) Master

   At this stage, the expert does not need principles anymore to guide their actions. Their actions are instantaneous, and little to no mental effort is required to keep track of their actions. Although, it is stated by Dreyfus [16] that "according to our model, there is no higher level of mental capacity than expertise." [16]

In an article by Benner [5], the Dreyfus model of classification was suggested to be applied to nurses to classify them by their proficiency levels. The authors conducted interviews and observations with fifty-one experienced nurse clinicians, eleven new graduate nurses, and five senior nursing students from six different hospitals to confirm their classification scale [5]. Five stages of proficiency were suggested [5]:

1) Novice

   The novice nurse is new to the field and will have no experience. They are taught about new situations in terms of "objective attributes" such as "weight, intake, temperature, blood pressure, pulse" [5]. The novice nurse is given rules to follow to guide their actions in practice. [5]

2) Advanced Beginner

   To become an advanced beginner, the nurse must demonstrate some 'acceptable' performance. For example, they are not as flustered during hectic situations, learning to manage the situation. At this level, the nurse has been able to cope with various situations known as 'aspects'. Lastly, the nurse has spent a significant amount of time with preceptors and new grads to learn how to recognize these 'aspects'. [5]

3) Competent

   The competent nurse will have two to three years of real work experience. Their actions will be thought of as a plan rather than a moment-to-moment coping with situations as they arise. These 'plans' can include 'objective attributes' and 'aspects' of the current (and possibly future) situations. [5]

4) Proficient

   The proficient nurse begins to see the situations they encounter instead of as individual events or aspects. Furtherly, the proficient nurse begins to understand how multiple aspects or attributes play a role in a situation. They also begin to recognize when a typical aspect is missing from a situation. Benner [5] states this as recognizing “when the normal is absent.” Decision-making, at this stage, is more intuitive than before as the nurse has been able to accumulate schemas of these various situations that they have encountered. [5]

5) Expert

   Lastly, the expert nurse does not rely on rules or guidelines to understand the situation at hand to determine the appropriate action. Their behavior at this stage is intuitive as their schema repertoire has grown significantly [5].

As mentioned above, there are concerns with high levels of cognitive load in health care as this can increase the likelihood of errors, which in the health care field can pose potentially fatal risks to the patients in the care of nurses. Therefore, such a scale is of interest to understand how much nurses can handle and which situations they can cope with on the job. Especially when considering new equipment featuring a different user interface that may work differently, such as infusion pumps.

Lastly, Renkl [45] states that many researchers have suggested more streamlined skill acquisition stages focusing on three main levels. [45] explained these three stages as the following:

1) Early

   In the early stages of skill acquisition, learners begin by understanding different aspects of the task at hand without seeing the larger picture or applying any learned knowledge since they have not learned any yet. [45]

2) Intermediate

   In the intermediate stage of skill acquisition, the learner begins to, and has a larger focus on, using what they have learned to solve problems within the task. [45]

3) Late

   Lastly, in the late stages of skill acquisition, the learner increases in speed and reduces errors or mistakes due to the learner's increased experience and practice. Lastly, the learner at this stage focuses on problem-solving (instead of reflections and self-explanations). [45]

These streamlined stages suggested by Renkl [45] show the same common steps of Benner's [5] and Dreyfus's [16] model of skill acquisition. In the first stage, the learner begins by learning and experiencing different aspects of the task. In the second stage, the learner starts to apply these aspects into some but smaller problem-solving tasks to begin to grow their repertoire of schemas. In the last stage, the learner has accumulated many schemas, and therefore, their behaviour becomes increasingly intuitive.

2.4.1 Supporting the Transition from Novice to Expert

To help support a novice learner through to the expert stages of understanding, it is essential to present the task in terms of aspects [16], also explained as isolated elements. Supporting learners with the theory of element interactivity is also known as intramodal improvement. As described by Cockburn [14], the users' practice in intramodal improvement focuses on one specific method and one specific aspect. An example given is as follows: "pointing and selecting with a mouse to select the bold function in Microsoft Word" [14]. Additionally, intermodal improvement [14] needs to be considered when supporting intermediate learners, whereby the learner can adapt to different, quicker methods to accomplish the same task. The example given by Cockburn is utilizing keyboard shortcuts instead of point and select. As learners approach the intermediate stage, it is vital to foster the extension of their vocabulary. Vocabulary extension is described by Cockburn as "ways to help users broaden their knowledge and their use of the range of functions available in an interface" [14]. Vocabulary extension is not learned automatically but through practice and the building of schemas, which introduces a fourth theory: task mapping [14]. Once the user has built a significant number of schemas and expanded their vocabulary accordingly, they will recognize the functions available, remember their intended use because of their stored schemas and decide on the appropriate action to complete their task. Once this has been achieved, the task has been successfully mapped; the user has been able to find a new way to achieve their goal while using multiple functions or aspects because of their available knowledge. Lastly, in the late stages, the user can use their schemas and mapped tasks to utilize them to solve new problems as they are encountered.

2.4.2 Cognitive Load Through the Transition

As explained above, intrinsic load gradually decreases as the user becomes more proficient in the skill or task at hand [42]. Once this occurs, memory resources are freed for germane and extraneous cognitive load to allow the formation and schemas and subsequently, for the user to increase in proficiency level [42]. It is most important for early and intermediate stage learners to optimize germane cognitive load to allow the formation of mental models to be able to remember what they need to do, how to accomplish it and move onto higher stages of learning [42]. The idea here is to reduce element interactivity to allow easy learning, lower intrinsic cognitive load, and free memory resources for germane cognitive load, allowing schemas to form [42]. The Fading Worked Out Solution [46], for example, states that by providing the user with a Worked-Out solution, and as they increase in competency, slowly “fade” out information from the solution. As they become more competent, more should be removed from the solution for the user to figure out on their own [46]. With this type of worked-out solution, the user must understand and work through a little more as they move forward. As their proficiency increases, the material's load is not significantly amplified due to their augmented understanding, which does not add more intrinsic cognitive load to the user, maintaining a good balance of all three types of cognitive loads [42]. On the other hand, expert users the most important thing, in this case, is to maintain consistency in intrinsic and extraneous cognitive load to avoid overloads (and errors) and work as efficiently and as quickly as possible with the prior knowledge they have. [42]

2.4.3 Design Guidelines

Many design guidelines, which are discussed in in-depth below, have been suggested to reduce cognitive load. The following three sections lists the bulk of them in relation to the type of cognitive load it supports. When we should implement these principles is a challenging question because it depends on user personalization. They are most important for novice users to, as previously mentioned, to foster the creation of schemas. These principles can also help expert users in new tasks; although, in an expert user where they already formed a schema, these principles can become detrimental and cause the expertise reversal effect.

2.4.4 Extraneous Cognitive Load

In 2010, it was suggested by Van Merriënboer [56] to use these design guidelines to manage extraneous cognitive load better:

1) Goal-Free Principle

The Goal-Free Principle entails changing a typical task that would ask for the goal directly (such as: Please list x reasons y) to a goal free task (such as: which are the reasons why …?). In a goal-free task the learner is encouraged to find the answer to the specified goal instead of guiding them down the path to the correct answer. [56]

2) Worked Example Principle

The Worked Example Principle suggests providing the learner with a complete solution to the task at hand. This will allow the learner to see what the outcome of the task should be and provide them with a version to look over. [56]

3) Completion Principle

The completion principle, much like the worked-out solution, is to provide a 'fill-in-the-blanks' type tasks where the learner must only figure out part of the solution and are helped through or given the rest. [56]

4) Split Attention Principle

The split attention principle suggests providing information or tasks from only one avenue versus multiple. This will diminish the learner's possibility of increasing cognitive load because they would be focusing on multiple places at once. [56]

5) Modality Principle

   The modality principle suggests replacing some text with a different modality, such as video or audio [56]. Since a person can handle processing multiple modalities such as visual and written simultaneously, presenting some text instructions in a different modality, can help reduce cognitive load rather than presenting everything via text. [32]

6) Redundancy Principle

   Similarly to the split attention principle, the idea behind the redundancy principle is to minimize the avenues of information coming in. Although in this case specifically, the ones that are explaining the same things, the redundant ones. [56]

2.4.5 Intrinsic Cognitive Load

In 2010, it was suggested by Van Merriënboer [56] to use these design guidelines for managing intrinsic cognitive load:

1) Simple to Complex Strategy

   The simple to complex strategy takes advantage of the isolated elements and interacting elements principles. At earlier stages of learning, present the tasks in an isolated elements fashion so the learner can understand each aspect independently. As their knowledge base grows, providing a more complex strategy, and introducing interacting elements to allow the learner to form schemas of the larger picture. [56]

2) Low-to-High Fidelity Strategy

   The low to high fidelity strategy suggests to begin by explaining a task in a low-fidelity environment such as text, video, or audio. As the learners' knowledge increases, change the presentation style to a high-fidelity environment such as a simulation or a Virtual Reality environment. [56]

2.4.6 Germane Cognitive Load

In 2010, it was suggested by Van Merriënboer [56] to use these design guidelines for managing germane cognitive load:

1) Variability Principle

   The variability principle suggests changing a task that describes one condition in multiple scenarios to describing one scenario from multiple conditions. [56] For example, if describing the different ways that a computer can fail; describe one way that a computer can fail on all types of operating systems. Then describe different types of failures within one operating system. This will allow the learner to understand these different "aspects" [16] individually before understanding multiple failures at once.

2) Contextual Interference Principle

   The contextual interference principle entails randomizing tasks versus presenting them in a natural order. This will allow the user to understand each task as an "aspect" [16] before understanding the task as a whole. At earlier stages of learning, the task may be overwhelming for learners. [56]

3) Self-Explanation Principle

   The self-explanation principle allows the user to explain that step of the task, when prompted, to reinforce their understanding of the task at hand. [56]

2.5 User Interface Modifications

Interface changes are inevitable. In the twenty-first century, our digital lives evolve so quickly to keep up the pace with our real lives. This entails that digital applications continue to change and transform themselves to serve us and our goals best. Although, these changes can quickly become unfavorable. This can be seen in an article by Cacitti [11], where they explored how accidents are derived from interface changes. It was found that negative transfer was most prevalent when participants were asked to use different keyboard keys to accomplish the same task, resulting in an increased number of errors [11]. This is assumed to be the case because it takes time, and experience as stated by Dreyfus [16], to become an "expert" and become familiar with these new keys. The authors concluded their study by suggesting that if changes are required on an interface, make more significant changes that are evident to the user, such as changing from keys to on-screen buttons, in order to reduce the possibility of negative transfer and errors in misleading schemas [11]. Although, this statement by Cacitti [11] is controversial. Others, such as Fleischmann [19] have found that users' prefer updates containing single features at them opposed to multiple features updated together.

This is only one of many examples of user interface modifications that could be implemented. Many exist, but they can all be categorized under these three types of interface changes:

- Aesthetical changes: colour change to buttons, text, background, foreground.
- Functional changes: a feature has been changed, added or removed.
- Input changes: input keys such as on-screen or physical inputs have changed.

2.5.1 Gestalt Principles

Gestalt principles are theories to explain how, we as humans, perceive and understand different visual aspects. While many Gestalt principles exist, over 114 stated by Graham [23], its' complete list would be too lengthy to report here. Therefore, the following fourteen were chosen to report as they are the most commonly found principles in recent literature [23] [12] [47] [38] [29] are summarized below:

1) Proximity [23] [12] [47] [38] [29]

Items visually situated in the same area will be perceived as a collection where items situated farther apart will not be. Subsequently, the farther items are apart, the more mental energy must be used to identify a link between the information if one exists. An example of this is shown in Figure 4. Since a column of links splits the paragraph, the reader must scan the (irrelevant) links to continue reading. According to the law of proximity, this column of links should be placed on the left or right [23].

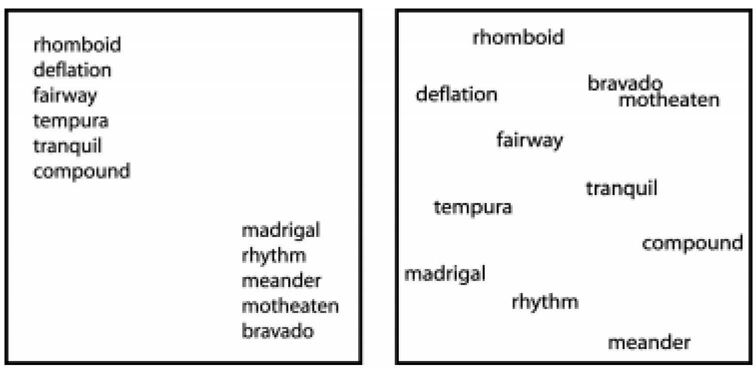


Figure 4 [23] – Law of Proximity

2) Common Direction/Fate [47] [38] [12] [29]

The law of Common Direction, also known as the law of Common Fate, states that objects that behave similarly, for example, flash simultaneously, change colours simultaneously, and are perceived as a group.

3) Simplicity [47]

The law of Simplicity states that items with simple designs are easier to see, understand and recognize.

4) Closure [23] [12] [29]

The law of Closure entails that, we as humans, can “close” an object, an audio or video clip, even if some of the information is missing for us. An example of this is the IBM logo in Figure 5. Even though a portion of the logo is not there, we can still distinguish three letters: I, B, and M.

Figure 5 [27] – Law of Closure

5) Continuation / Continuity [23] [12] [29]

Similarly, to the law of Closure, the law of Continuation states that when a broken line or choppy audio clip is presented, we as humans can fill in the gaps and perceive the item presented to us. For a song, it will not be perceived as a different song, we can sing along and continue the song regardless of the break. Similarly, to a dotted line will be perceived as a full line as can be seen in Figure 6.

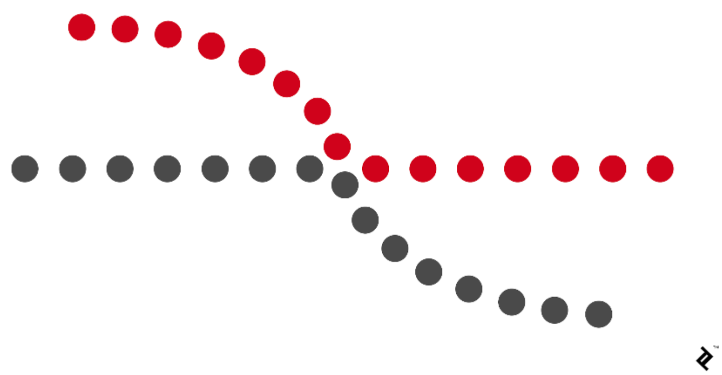


Figure 6 [54] - Law of Continuation / Continuity

A second example is this snapshot from Sprig in Figure 7, which demonstrates a three-step process to use their app. Regardless of whether the three steps or circles are separated by white space, we can still recognize that these three steps are one instruction and belong in one group together.

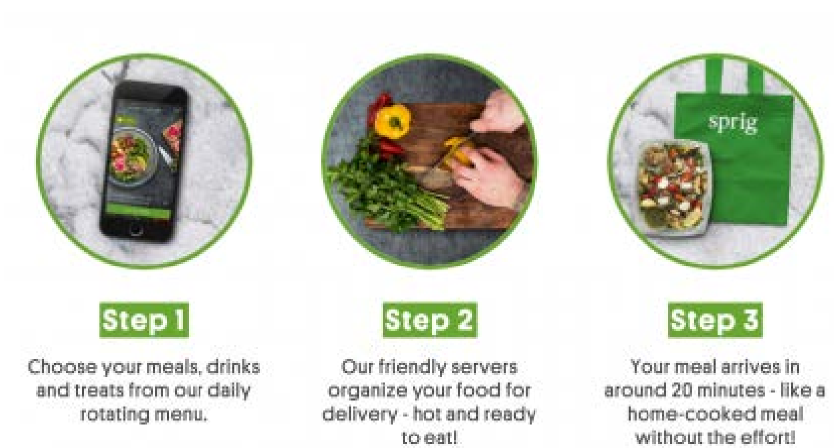


Figure 7 [55] – Law of Continuation / Continuity

6) Similarity [23] [12] [47] [29]

The law of similarity explains that are visually alike are perceived as being grouped together, where dissimilar items are seen as separately, regardless of the distance between them. An example is given, again with links in Figure 8, where links that are the same colour seem to be part of the same group.

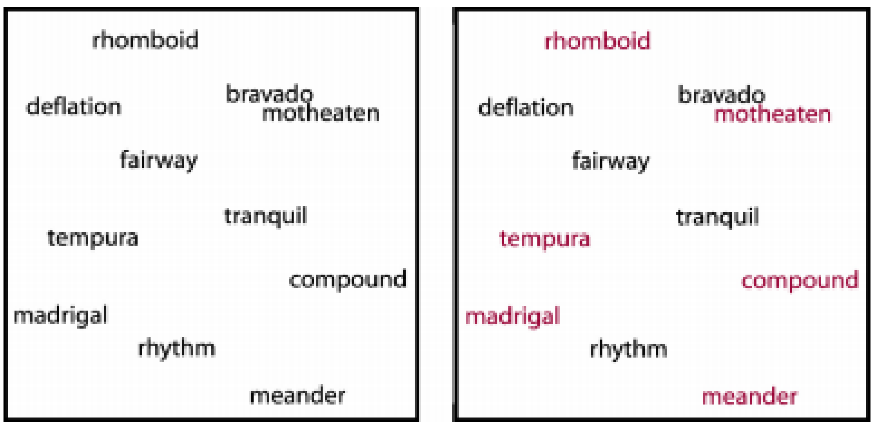


Figure 8 [23] – Law of Similarity

7) Focal point [12] [47]

The law of Focal Point entails that an item with a very distinct colour (such as a neon colour, a distinguishable colour or an animation) will create a fixation point for the user where all their focus will be drawn to that one item. An example can be seen in the following figure, Figure 9.

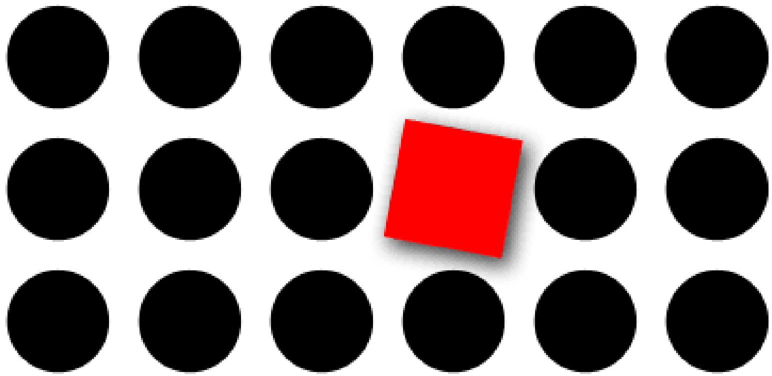


Figure 9 [55] – Law of Focal Point

8) Figure-ground [23] [12] [29]

The idea behind this principle is that having the foreground and background colour as two distinct colours allows the viewer to distinguish between them. When these colours are too similar, they are harder to tell apart, therefore, being more difficult to read or see. An example can be seen in Figure 10.

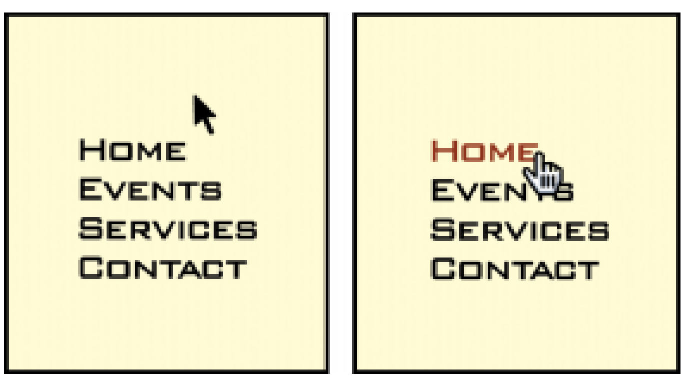


Figure 10 [23] – Law of Figure-Ground

9) Belongingness [12]

The law of Belongingness explains that an item is perceived to only belong to one group unless separated by another item. An example of this is with 'beeps' for notifications. If a 1ms beep followed by a 1ms pause is used to indicate Warning A, but then a 2ms beep followed by a 1ms pause is heard, these two beeps will be perceived as two different warnings.

10) Good Figure [54] [29]

The law of Good Figure states that when we are presented with a picture of overlapping items, we will interpret this combinational figure as each individual shape. For example, in Figure 11, the Olympics logo will be interpreted as five overlapping rings instead of an ambiguous configuration of parts of circles.

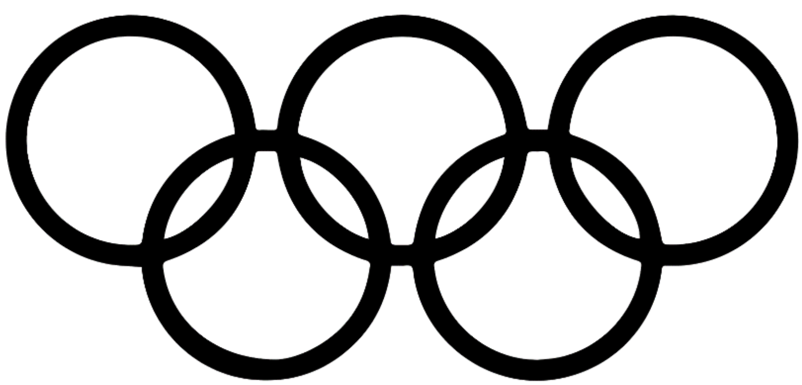


Figure 11 [54] – Law of Good Figure

11) Common Region [38]

The principle of Common Region explains that items that are englobed by some delimiter, such as an outline (square, circle) as in a table, are interpreted to be a group. This can be seen in Figure 12.

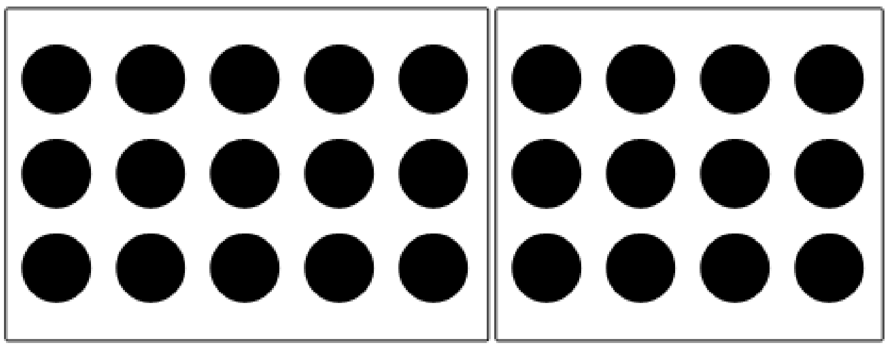


Figure 12 [55] – Law of Common Region

12) Symmetry [12]

The law of Symmetry states that it is typically easier to perceive and understand items that are balanced and symmetrical. An example being Figure 13.

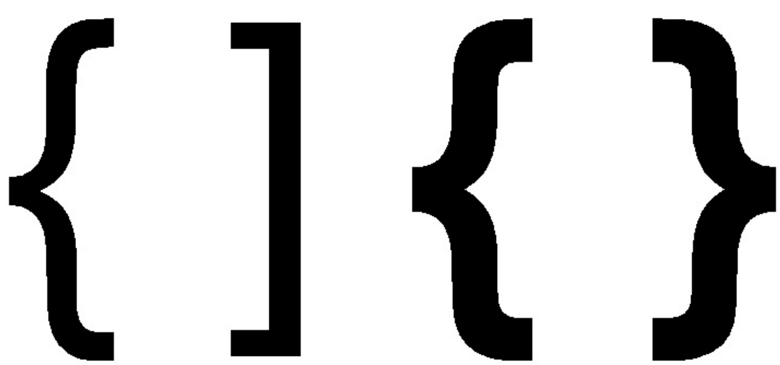


Figure 13 [10] – Law of Symmetry

2.5.2 Visual Search on Web Pages

In a series of articles by Jonathan Ling and Paul van Schaik, in the early 2000s, multiple theories surrounding various visual modifications affecting visual search on web pages were tested. These articles reviewed the effects of line spacing and text alignment [35], font type and line length [34] and text and background colour [33].

Based on the results from the [33] study, it was deduced that colour combinations do affect accuracy and speed [33]. Of all the colour combinations tested, the Green/Red combination performed the least well. Although, the default combination of Blue/White was the most preferred [33].

From the results of the [34] study, it was deduced that long line lengths should be used when the intention is to scan the information [34]. Shorter line lengths should be used when the information is to be read thoroughly [34]. As well as Arial should be used over Times New Roman [34].

From the results of the [35] study, it was deduced that performance was optimal with double line spacing versus single or single and a half spacing [35]. Similarly, the performance scores were higher with left-aligned text versus justified text [35]. It was stated that justified, single-spaced was more challenging to search, although increasing the line spacing to double-spaced while keeping the justified text removed this difficulty [35]. Despite the performance scores showing these results, the subjective assessments showed that participants preferred justified text to left-aligned [35].

2.5.3 Keyboard Layouts

To further investigate the effects of interface modifications, research has been covered surrounding the idea of changing the keyboard layout of a device. In this case, the method in which the user is inputting information may be slightly different and may cause some differences in performance and efficiency of entry.

An article by Jokinen [30] aimed to understand the visual search and learning of new keyboard layouts. The authors' goal was to create a computer simulation model to simulate a user's learning of a given keyboard. A data set was needed to create and test the model. Therefore, two studies were conducted to gather sufficient data to train the models. The first study aimed to gather data to represent how a user learns an entirely new layout.

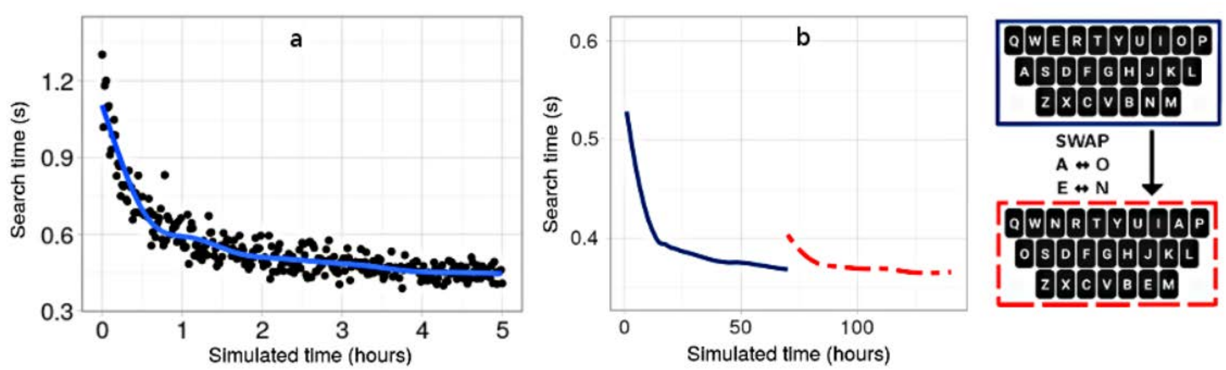


Figure 14 [30] – Simulation Results

The second study aimed to gather data to represent how a user adapts to a changed layout . Their model simulated two keyboards: the QWERTY and a 'swapped' layout of a QWERTY-like keyboard where the A key was changed for the O and the E key was changed for the N. The simulation results are as follows in Figure 14. [30]

The interesting result with this simulation is that after the first half of the simulation, the search time begins to steady for the remaining time of the simulation. This shows that there is a learning curve at the beginning but after some time, the layout is learned [30]. Similarly, to when the swapped key is introduced, the results are higher for some time but then steady out once it is 'learned' [30]. Similar results to the ones found from these simulations can also be seen in other articles, including [6] [39] and [17].

Firstly, in an article by Bi [6], their goal was to find a keyboard layout that would ultimately reduce the number of errors that occur, increase speed, and that is easy to learn when used with gestures [6].

This experiment showed that regardless of the mean input time being lowest during the first few repetitions on the qwerty keyboard, the qwerty layout and its' variations balanced out after multiple repetitions [6]. This can be seen in Figure 15 and 16. The variations were even found to have a shorter input time after the third repetition, similarly to the study by Jokinen, where after the first few simulation hours, the results balanced out for any keyboard variations.

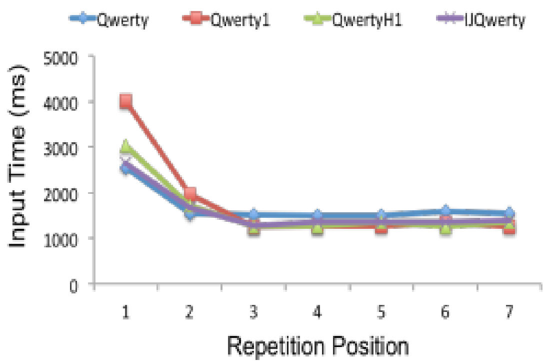


Figure 15 [6] – Steady Results

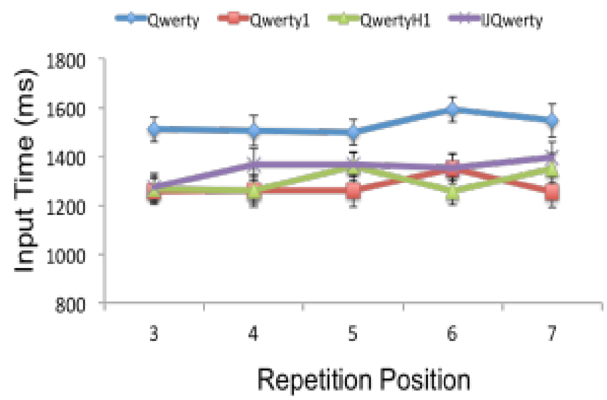


Figure 16 [6] – Steady Results in Final Five Repetitions

Secondly, in an article by Niemela [39], four visual layout conditions were used for these experiments being: grouped and ungrouped, coherent, and non-coherent interface elements. In a series of experiments, the participants were asked to perform a recall task of either the label or the label's location [39]. From these experiments, similar results as the previous two were seen.

Thirdly, in an article from Dunlop [17], a sath keyboard optimization is explored in a user study. The results of this task showed similar results that the first few days were low, but as time increased, the user's speed also increased, becoming closer to the qwerty performance score. The results of the NASA-TLX surveys also followed this pattern as well.

As shown in these studies, when a user is encountered with something different than expected, whether it be a change in a keyboard or a change in layout, they will adapt and return to their typical performance scores with time. In the first few trials, the known condition may be the highest. After these initial trials, the results, whether performance scores or time taken depending on the study, they flatten out for all conditions, even the new ones.

Chapter 3 – Experiment

The purpose of this experiment is to begin to address to what degree user interface changes affect a user. Therefore, for this experiment, the goal is to test small interface changes to better understand to what small extent the users are affected. Once the feasibility of small changes is evaluated, further work could include testing more significant changes to better understand them as well.

3.1 Experimental Design

An experiment was designed that consisted of a game similar to the Simon game. Figure 17 displays the game box of this game. Simon is a memory game in which a series of lights flashes in sequential order, and the player is required to remember and enter the lights in the order they appeared. The game design utilizes shapes instead of colours. This experiment aims to better understand how small changes to user interface input, such as those applied, affect the user. These results can be used as a limit to understand if, more or less, larger or smaller changes can be applied without imposing any effects on the user in question.

Figure 17 [53] – The Simon Memory Game

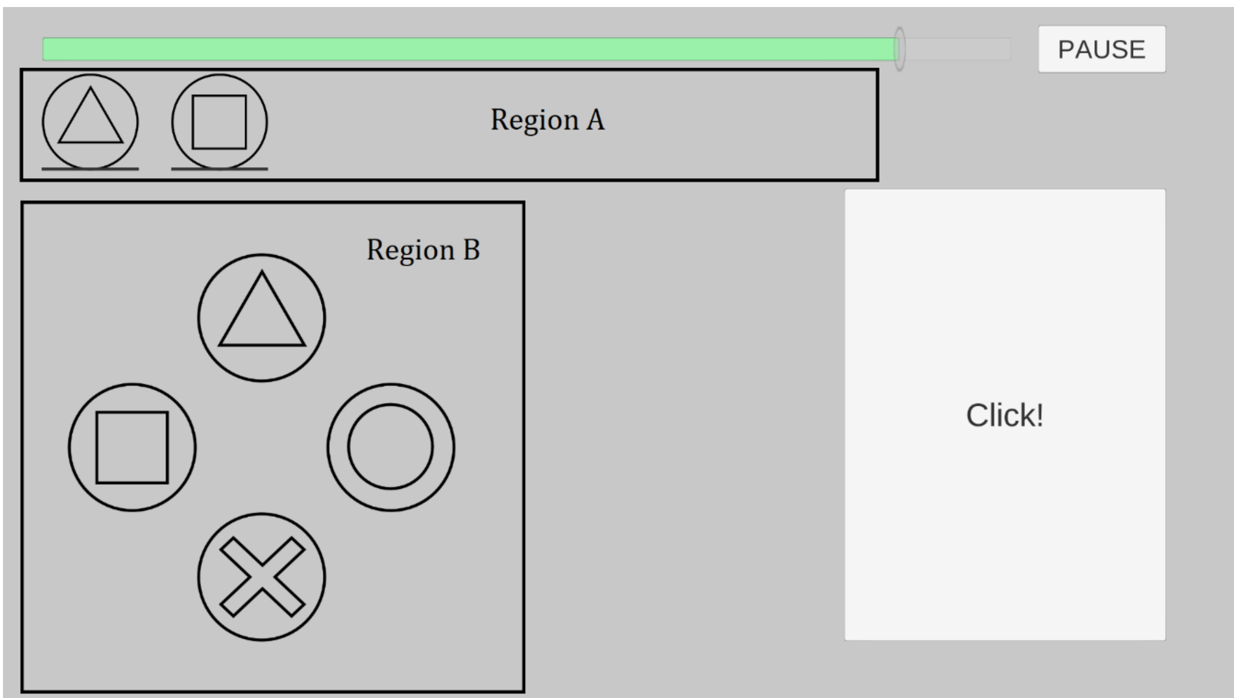


Figure 18 – Interface Design

3.2 Interface Design

The interface design is shown in Figure 18. The sequences to enter would appear in Region A. The sequence would also appear in a sequential and levelled order where the participant would receive the first shape in the sequence; once entered correctly, then two shapes would appear, then three, so on, up to seven. This appearing order was chosen for many reasons. Firstly, to present an increasing challenge to the player, a few shapes may not be difficult to remember, but once this number increases, the challenge also increases. Secondly, to mimic real-world activities. Our daily activities continuously compound in our everyday lives, adding more and more to our plate, adding more things to remember as we go on. Lastly, this was done to accommodate Miller's theory of being able to remember 7 +-2 chunks of information at any given time [20]. The inputs in Region B are used to enter the sequence shown. Shapes were chosen as they are understandable across many languages; regardless of the shape's name, a shape can still be understood and interpreted as the same. The inputs were chosen to be englobed within an outlining border (circle) for multiple reasons. Firstly, to present a delimiter to the button. For the purpose that participants would not have to spend time trying to click within the shape for the game to successfully detect an input. The purpose of the placeholder is also to differentiate the four inputs. Similarly, the shapes are presented in a black/dark tone present no difficulties to individuals with colour vision deficiency, increasing the target population to recruit. The green bar along the top indicates the amount of time remaining to enter in the sequence to introduce time pressure, stress and a sense of urgency in the player. Given that the experiment needed to be converted to an online remote delivery format, a pause button was included if other household distractions were to occur to minimize the number of flaws in the data due to external distractions. Lastly, the white "Click!" box on the right-hand side is used as a Dual-Task to track cognitive load. This button would appear every seven seconds. Through pilot testing, this was found to be the optimal timeout period between box appearances where it was not too often, nor be too distracting but often enough to consistently keep the user attentive to the button’s appearance. Once clicked, a reaction would be tracked in the database. This data is then used to determine the participant's reaction time to the stimulus which is used to infer the user’s cognitive load.

3.3 Hypothesis

This type of game (memory game) was chosen to test how Gestalt Principles affect the user when the user’s attention is focused on remembering the sequence presented to them, especially when the items' length is continuously changing.

The following research questions are posed:

RQ1: Does following the Gestalt law of Figure-Ground, Proximity or Closure increase cognitive load, speed, and performance?

RQ2: Does compounding different changes/principles affect cognitive load, speed, and performance?

RQ3: Does changing the placeholders affect a user's mental model, given that they remembered the original placeholders by their position and not their label?

RQ4: Do participants encode the sequence differently than they are presented? Does encoding in this fashion affect performance, speed, and cognitive load?

RQ5: Does muscle memory come into play when trying to remember a sequence of steps to complete a task?

3.4 Classes of Changes

To find which Gestalt principle variation to test, four classes (see below) were created, where each class groups changes following one central principle relevant to the game's interface design and a list of all possible changes that could be made onto the interface from that principle. The following three Gestalt principles were chosen to test as they are the three most generic and unique laws; out of all Gestalt principles, some laws such as Law of Continuation is derived from the Law of Closure, this is only one such example. In the list, the numbers within parentheses represent the number of possible changes that could have been made for that change, followed by all or a few examples of the changes that could have been made. Below each of the three laws, the various changes that could be implemented from this law are listed. Beside this change, the number of variations for that change is stated with a number in parentheses; as well, each variation is listed underneath.

1. Changes testing the Law of Figure-Ground

   The Law of Figure-Ground would be implemented by changing the weight and colour fill of the shapes, borders and a combination of both.

   a. Shape and Border Weight (3)

      i. (1) 1pt (2) 2pt (3) 3pt

   b. Filled shapes (3)

      i. (1) light grey (2) dark gray (3) black

   c. Filled border (3)

i. (1) light grey (2) dark gray (3) black

2. Changes testing the Law of Proximity

The Law of Proximity would be implemented by changing the size of the shape (but not the border), shape and border and the spacing between the inputs (as for them to be further apart).

a. Size of shape and border (3)

i. (1) 25% smaller (2) 50% smaller (3) 75% smaller

b. Size of the shape in relation to the border (3)

i. (1) shape 25% smaller (2) shape 50% smaller (3) shape 75% smaller

c. Spacing (3)

i. (1) 25% separated (2) 50% separated (3) 75% separated

3. Changes testing the Law of Closure

The Law of Closure would be implemented by changing the style of the shape within the border (but not the border), the style of the border (but not the shape) and the overall configuration of the inputs (instead of configured in a diamond shape, similarly to the configuration of a gaming controller, in another configuration such as a square).

a. Shape Style (6)

   i. (1) dotted (2) squiggly (3) double border….

b. Border Style (6)

   i. (1) dotted (2) squiggly (3) double border….

c. Overall configuration of inputs (3)

   i. (1) square (2) diamond (3) no particular shape distinguishable

4. Changes testing memory

The memory change has the purpose of testing how different characteristics of the buttons affects the users' capacity to memorize the input button's locations, label and the sequence.

a. Placeholder of Shape (4)

   i. (1) Shapes: (square, triangle, circle, x) (2) More Shapes: (pentagon, hexagon, star, diamond) (3) letters (4) numbers.

b. Placeholder of Border Surrounding Shape (4)

i. (1) Shapes: (square, triangle, circle, x) (2) More Shapes: (pentagon, hexagon, star, diamond) (3) letters (4) numbers.

c. Position of inputs (3)

i. (1) shift shapes counter clockwise (2) shift shapes clockwise (3) swap with the shape it is facing (top with bottom and left with right)

3.4.1 Evidence to Support Changes

Evidence to support the Class 1 changes can be found in [33]. Findings from this article suggest that text-background colour does affect performance in information retrieval speed. In our case, given that we want to eliminate colour, shading could be tested in its' place, and therefore, the shading that is easiest to see should increase the user's speed. Furthermore, according to Plouznikoff [43], highlighting the needed buttons, in this case, with the darkest fill, increases user speed. Lastly, according to the study done by Moller [38], acceptance is greater if Gestalt principles are implemented, meaning that little disruption should be caused if the gestalt principles are followed.

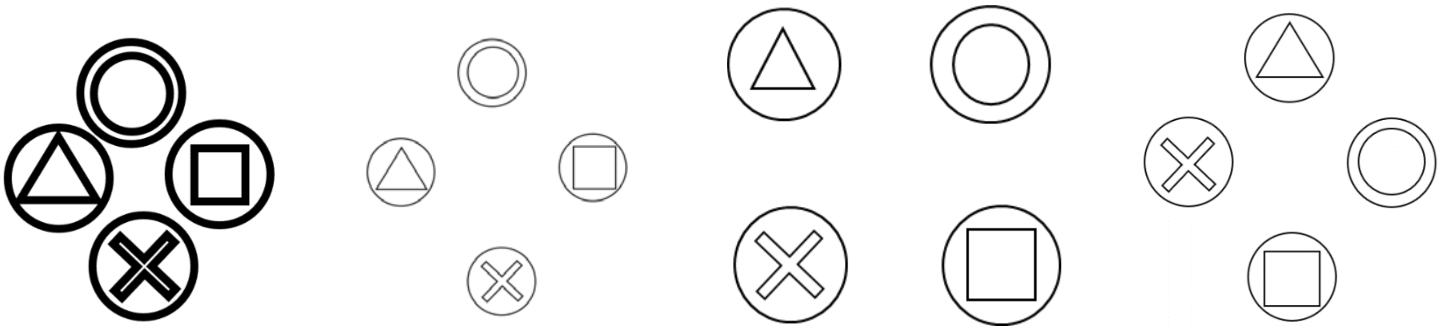
Evidence to support the Class 2 changes can be found in [39]. Findings from this article suggest that coherency enhances label recall. Given this, if coherency is held, regardless of the spacing between the icons, it should not disrupt the user. Furthermore, as can be found in [34], using the standard font of Arial, a "normal" or a preferred size should present the least amount of disruption. Lastly, according to the study done by Moller [38], acceptance is greater if Gestalt principles are implemented, meaning that little disruption should be caused if the gestalt principles are followed. We should expect an increase in speed [43] and minimized errors when this law is maintained, and the user's preference is achieved [34]. For some people, larger buttons may be helpful. For others, smaller buttons may be beneficial, which may affect their performance in the task.

Evidence to support the class 3 changes can be found in [34], using the standard font of Arial, a "normal" size should present the least amount of disruption. Lastly, according to the study done by Moller [38], acceptance is greater if Gestalt principles are implemented, meaning that little disruption should be caused if the gestalt principles are followed.

Evidence to support Class 4 changes can be in the study done by Moller [38]; acceptance is greater if Gestalt principles are implemented, meaning that little disruption should be caused if the gestalt principles are followed.

3.4.2 Changes Chosen

From this list, four changes were selected to test is the following:



Class 1 – Shape and Border Weight

Class 2 - Spacing

Class 3 – Overall configuration of inputs

Class 4 – Position of inputs

Figure 19 – Chosen Changes

These were selected from the list for the following reasons. To begin, only one change per class and one variation of that change could be included in order to decrease the number of groups for the experiment and, therefore, the number of participants required. Firstly, for class 1, the shape and border weight change were selected over the shape or border fill because the fills may not affect the user as much since the input changes will not be as evident due to the small screen size of a mobile device. Secondly, the spacing change was chosen from class 2 because the remaining changes (size of shape in relation to border and the size of border and shape) given the mobile devices small screen, there was not much screen real estate to augment or decrease the size of the input, the only feasible option seemed to be the spacing change. Thirdly, shape style and border style were also suspected to be less evident to the user on a small screen. Therefore, the overall shape position was chosen. Lastly, for class 4, the position change was chosen as, unlike the other changes, this would be testing the user's memory the most. This is because the shapes themselves will be moving, also implicitly changing the pattern. With this, we will see more errors compared to the shape and border placeholder if the user's mental model or muscle memory pattern is disrupted.

3.5 Procedure

The experiment began with and explanation as to what is expected from the participant for the duration of the experiment such as, how to play the game and an overview of the consent form, the explanation provided can be found in Appendix E. Participants were then asked to sign and date the consent form; the consent form provided can be found in Appendix D. Followed by a demographic questionnaire which can be found in Appendix A. Once filled out and signed, the participants were instructed to play a test game to understand and become comfortable playing the game. Once they played for as long as they wanted, they switched to the experimental trials. In these trials, they played a total of 14 games. Games 1,3,5,7,9,11, and 13 were the original interface, as shown in the table 1 below. Games 2,4,6, and 8 introduced a change, and games 10, 12 and 14 introduced a combination of changes. This can be seen in the table 1 "Order of the Experiment" below. A view of the game interface and a series of pictures which demonstrates the various steps of a typical game can be found in Appendix H. After game 8, the participants were instructed to fill out a NASA-TLX rating scale before continuing onto the final six games. After all, games were completed, the participants were asked to fill out another NASA-TLX rating scale, and the weighting phase was completed with the participant. The forms used can be found in Appendix C, was also completed with the participants and the final questionnaire, can be found in Appendix B.

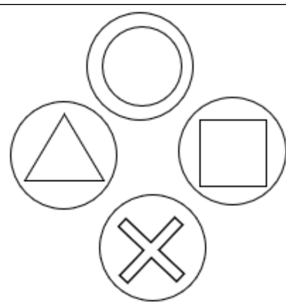
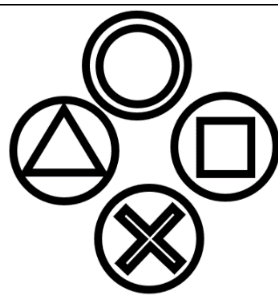
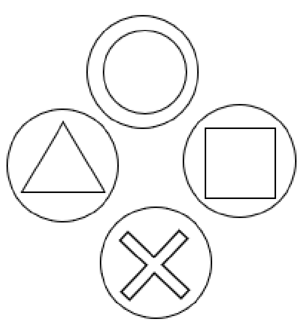
3.5.1 Order of the Experiment

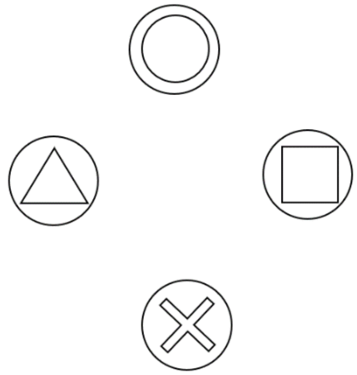
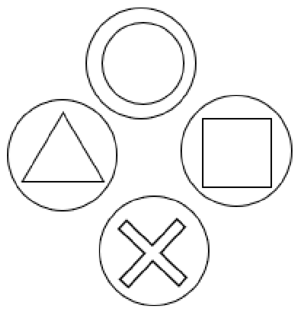
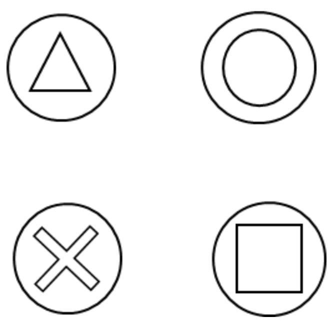
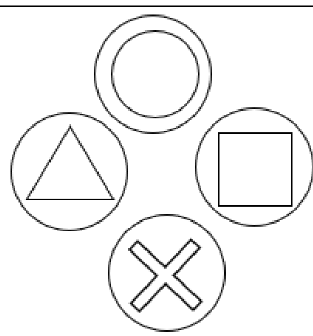
The experiment games were conducted in the following order with each of the participants. Depending on the counterbalance group that they were assigned to (listed in the fourth column from the right of the following table), the users receive the game groups in a different order. Regardless of the counterbalance group that they were assigned to, the game sets E, F and G (the combinational change games) were experienced in the same order.

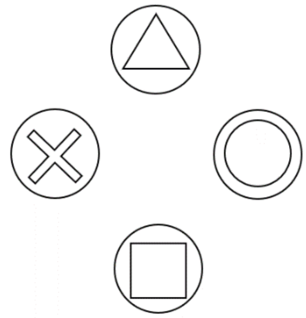
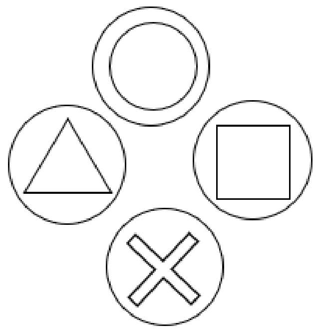
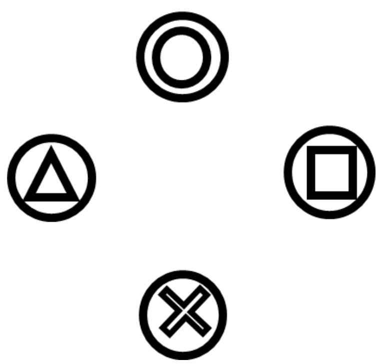
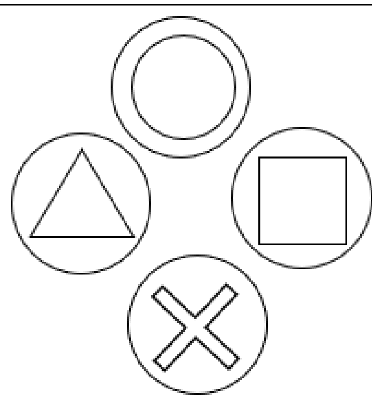
Counterbalancing was implemented to ensure that effects generated from the changes were not directly caused by the order in which they were presented [44]. Therefore, by creating these counterbalance groups, if an effect was found, it could be ensured it was not caused by the order in which it was presented.

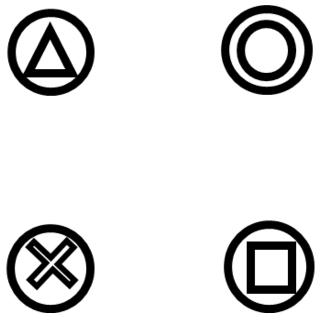
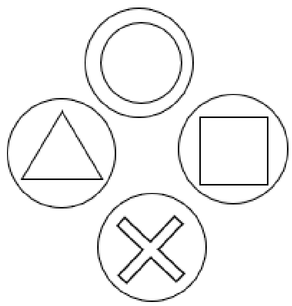
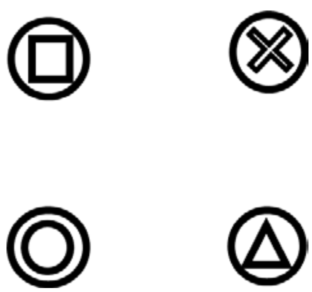
The counter balance groups are as follows, as included in Table 1: Group 1: ABCD, Group 2: ABDC, Group 3: ADBC, Group 4: DABC, Group 5: ACBD, Group 6: ACDB, Group 7: ADCB, Group 8: DACB, Group 9: CABD, Group 10: CADB, Group 11: CDAB, Group 12: DCAB, Group 13: BACD, Group 14: BADC, Group 15: BDAC, Group 16: DBAC, Group 17: BCAD, Group 18: BCDA, Group 19: BDCA, Group 20: DBCA, Group 21: CBAD, Group 22: CBDA, Group 23: CDBA and Group 24: DCBA. Therefore, if a participant is assigned to counterbalance group 5, they would receive game pair A (change of law of figure-ground), game pair C (change of law of closure), game pair C (change of law of proximity) and then game D (change of memory), then the combinational changes in order.

Table 1 – Order of the Experiment

Game Groups	Trial #	Button Style Presented	Counterbalance groups	Law Tested
A	1		ABCD, ABDC, ADBC, DABC, ACBD, ACDB, ADCB, DACB, CABD, CADB, CDAC, DCAB, BACD, BADC, BDAC, DBAC, BCAD, BCDA, BDCA, DBCA, CBAD, CBDA, CDBA, DCBA	None
	2			Law of Figure-Ground
B	3			None

	4			Law of Proximity
C	5			None
	6			Law of Closure
D	7			None

	8			Memory
	9			None
E	10			Law of Figure-Ground and Proximity
F	11			None

	12			Law of Figure-Ground, Proximity and Closure
G	13			None
	14			Law of Figure-Ground, Proximity, Closure and Memory

3.6 Participants

A total of 32 participants were recruited from Laurentian University from the Faculty of Health (13), Faculty of Science, Engineering and Architecture (16), Faculty of Arts (2) and Faculty of Management (1). Seventy-five percent of the participants were between 18 and 25; the remaining twenty-five percent were between 26-39. About forty percent of the participants use 1-2 devices daily, about half use 3-4 devices daily, and the remaining participants' (5) use five or more daily. All participants' use a cellphone, laptop or desktop daily; a third use either a game console or wearable every day. Sixty percent considered themselves 5/5 proficient using their cellphone; the remaining scored themselves 3/5 or 4/5. A third rated themselves 5/5, a third 4/5, and the rest 3/5 or 3/5 when considering their proficiency on a laptop or desktop. Just over half of the participants use their cellphones on an active basis for 1-3 hours per day, a third use their laptops on an active basis for 4-5 hours per day, another third for 8+ hours and the remaining 1-3 or 6-7 hours per day. More than eighty percent of participants considered themselves 4/5 or 5/5 on the input tasks on a keyboard, input on a touchscreen and searching on the web. Lastly, 21 participants stated to prefer a mouse, keyboard and trackpad combo, nine preferred touchscreens, and 2 stated no preference. Thirty participants used their mobile devices; one used a tablet and one used a laptop due to technical issues.

Given this variability in participant background experience and proficiency level, it is possible that some participants may have been at an advantage or disadvantage. Although, given that participants were given as much time as needed to practice on a test trial before beginning the experimental trials and that some participants played for longer and some for shorter lengths of time, I assume that participants' who needed more practice, played in the test trial longer, where the experienced players, played for a very short amount of time. Therefore, it is possible some participants were at an advantage or disadvantage, but the implementation of a test trial was included in order to reduce this.

There may implicitly be a bias given that seventy-five percent of participants were between the ages of 18 and 25 and the remaining between the ages 26 to 39. This indirectly excludes the elderly population.

3.7 Data Collection

Data was collected in four ways. Firstly, by using a dual task. This task is shown as a secondary stimulus task. The participant is required to periodically respond to the stimulus by clicking the "stimulus" button. The difference in time between the button appearing and the stimulus's response will be tracked as reaction time to infer cognitive load. The dual-task paradigm was chosen as an alternative method to collecting cognitive load data in place of an objective measuring tool and with a purpose of validating the NASA-TLX adjusted rating scores due to its' implicit subjectiveness. Secondly, time to complete a trial was collected between the start of the sequence and the end of the sequence input, recording both failed and successful entries to infer the input speed. Thirdly, correct and incorrect inputs were collected to infer input accuracy and performance. Finally, the participants were administered three questionnaires: a demographics questionnaire before the experiment, a post-experiment/subjective questionnaire after all trials are completed, and one NASA-TLX after the counterbalanced trials, and another after the compounded trials.

In a summary article by Alpert [1] where articles were reviewed by the popularity of choice of subjective cognitive load rating scales, out of 34 articles reviewed, 80% of the articles used the NASA-TLX scale to evaluate cognitive load or to compare against a new scale to confirm its' validity.

The dual task was utilized in place of an EEG to collect quantifiable cognitive load data due to limitations with testing in-person because of COVID-19.

Chapter 4 – Results

In this chapter, the results from the experiment conducted are summarized and analyzed. This includes: the statistical analyses of the performance scores, namely: cognitive load, time and errors, the analysis of the NASA-TLX surveys for cognitive load data and lastly, the analysis of the post-experiment questionnaire. A general summary of all the results can be found in the following figure (Figure 20) and following table (Table 2). Individual figures for each variable, time, errors and cognitive load, can be found in Figure 21, 22 and 23, respectively.

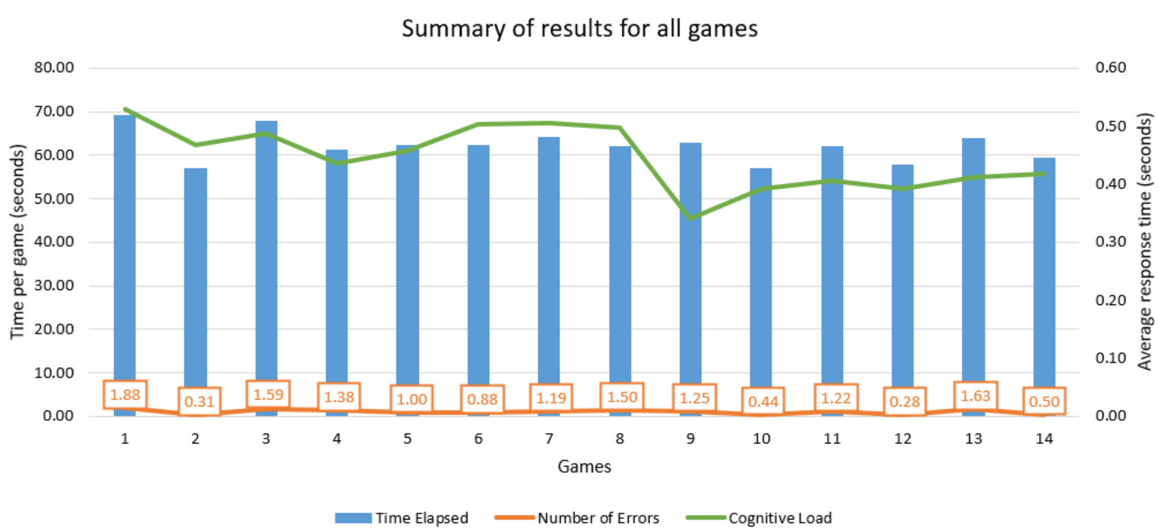


Figure 20 – Summary of Results

Table 2 - Results

	A		B		C		D		E		F		G	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Time	69.13791	57.0808	67.9235	61.2731	62.26981	62.3569	64.31274	62.1929	62.75412	56.91766	62.19708	57.8334	63.82919	59.45232
Errors	1.875	0.3125	1.59375	1.375	1	0.875	1.1875	1.5	1.25	0.4375	1.21875	0.28125	1.625	0.5
CL	0.529603	0.468215	0.486342	0.435474	0.457358	0.502876	0.50546	0.498077	0.340193	0.392252	0.406022	0.391922	0.412141	0.41843

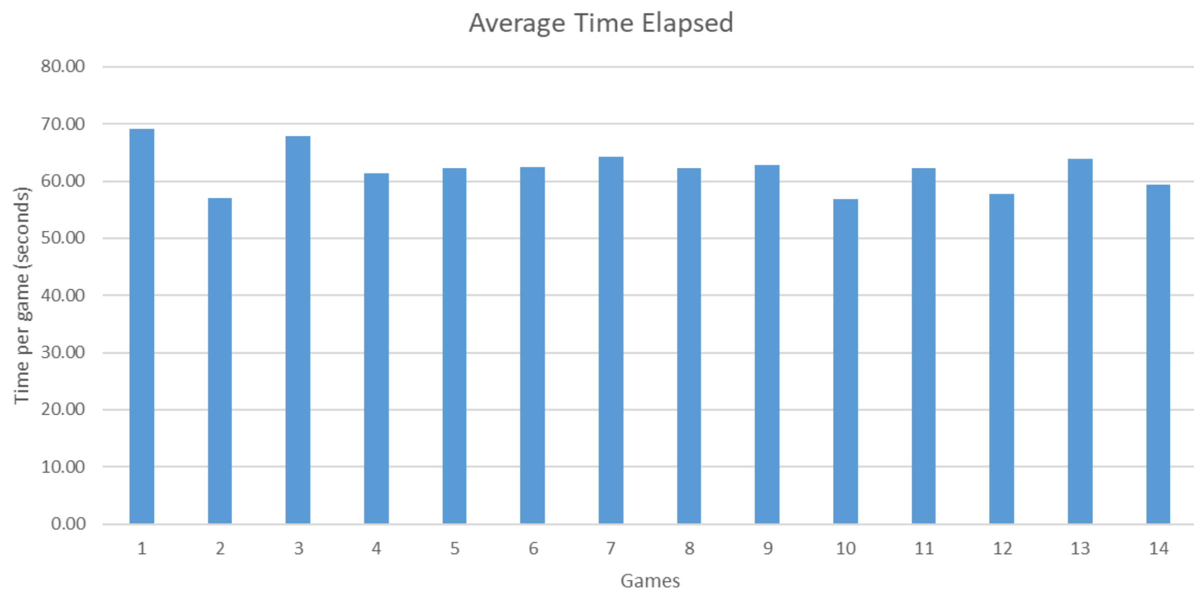


Figure 21 – Average Time Elapsed Per Game

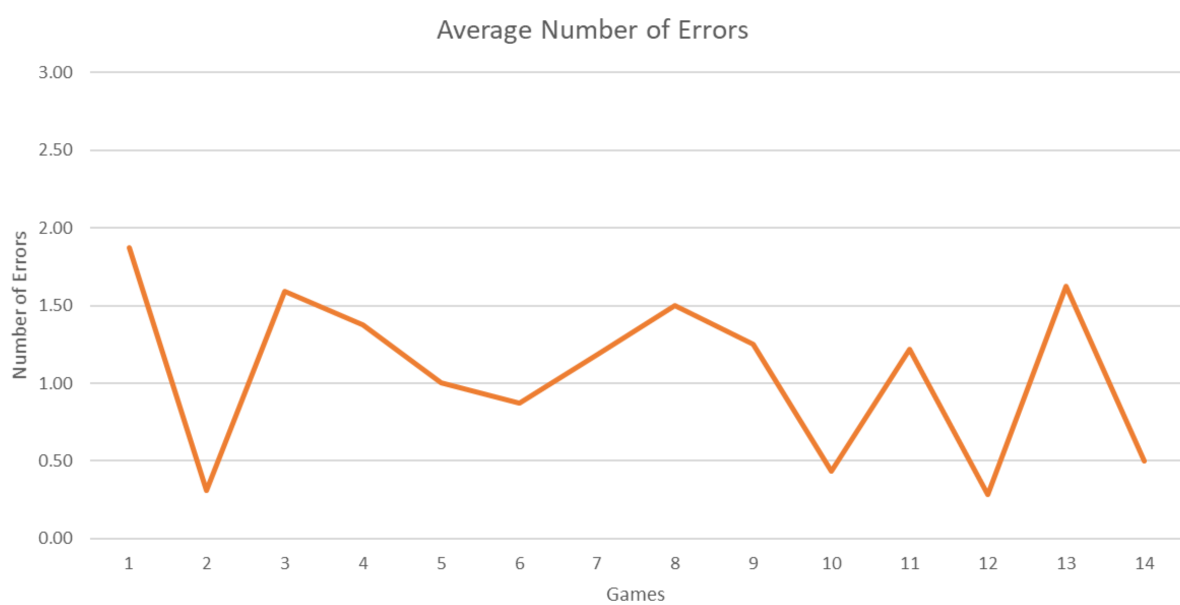


Figure 22 – Average Number of Errors Per Game

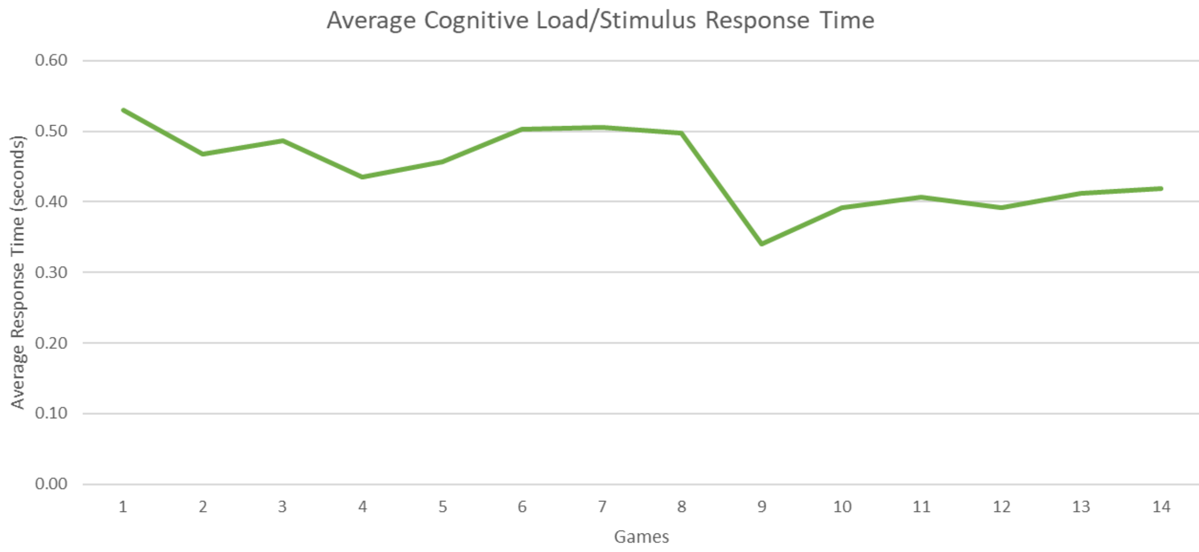


Figure 23 – Average Stimulus Response Time Per Game

In general, every game pair, for the most part, has a lower completion time, number of errors, and cognitive load response time than the preceding game. The exceptions to this are as follows. For time, game set C was ~0.1 seconds higher. For errors, game set D was 0.3125 higher. Lastly, for cognitive load response time, game sets C, E and G were exceptions with a difference of ~0.5 seconds, ~0.5 seconds and ~0.06 seconds, respectively.

In an overall analysis of the average scores for all participants, there were no unexpected correlations between any variables. The game with the longest completion time, was also the game with the highest number of errors and cognitive load response time other than the exceptions presented above. In terms of an individual participant analysis, this same trend followed.

4.1 Statistical Analyses

A multitude of t-tests and ANOVA tests were conducted testing each possible pair of games (comparison of 2 games, 3 games, 4 games, etc..) for a total of about 40 000 tests. These tests were conducted with a confidence interval of 95%. Therefore, any p-value returned below 5% would be rejected as “not statistically significant”. These tests were conducted to compare several independent means [18] and with this, to identify statistical significance between them. From these, tests relevant to the three variables of interest (errors, time, and cognitive load) will be reported in their respective sections below. A full list of all tests could not be included due to its’ sheer size. Post-hoc tests of relevant ANOVA tests were also conducted for every combination of more than two games that showed statistical significance.

4.1.2 Time

Table 3 – Time Pair Analyses

Game	1	2	3	4	5	6	7	8	9	10	11	12	13	14
p-value	0.001		0.006	0.136	0.981		0.009	0.019	0.089	0.017	0.088	0.032		0.000
Game Set Combination	(1,2)		(2,3)	(3, 4)	(5, 6)		(2,7)	(2,8)	(9,10)	(10,13)	(11,12)	(12,13)		(1,2,14)

For game set A, B, D, E, F and G, this variable averaged a lower time for the changed game than the original game. The only set that was an exception to this is game set C. The following results discussed can be found in Table 3.

To show statistical significance for game 1, games 1 and 2 are compared. The results of this t-test analysis show a p-value < 0.001 showing statistical significance. To show statistical significance for game 3, games 2 and 3 are compared. The results of this t-test analysis show a p-value < 0.006 showing statistical significance. To show statistical significance for game 4, games 3 and 4 are compared. The results of this t-test analysis show a p-value < 0.136 showing that it is not statistically significant. No other combination of games, including the 4th game, was found to be statistically significant. To show statistical significance for game 5 and 6, these games are compared. The results of this t-test analysis show a p-value < 0.981 showing that it is not statistically significant. No other combination of games, including the 5th and sixth game, was found to be statistically significant. To show statistical significance for game 7, games 2 and 7 are compared. The results of this t-test analysis show a p-value < 0.009 showing statistical significance. To show statistical significance for game 8, games 2 and 8 are compared. The results of this t-test analysis show a p-value < 0.019 showing statistical significance. To show statistical significance for game 9, games 9 and 10 are compared. The results of this t-test analysis show a p-value < 0.089 showing that these two games are not statistically significant. No other combination of games, including the 9th game, was found to be statistically significant. To show statistical significance for game 10, games 10 and 13 are compared. The results of this t-test analysis show a p-value < 0.017 showing statistical significance. To show statistical significance for game 11, games 11 and 12 are compared. The results of this t-test analysis show a p-value < 0.088 showing that they are not statistically significant. No other combination of games, including the 11th or 12th game, was found to be statistically significant. To show statistical significance for game 12 and 13, these two games are compared. The results of this t-test analysis show a p-value < 0.032 showing that they are statistically significant. To show statistical significance for game 12 and 13, these two games are compared. The results of this t-test analysis show a p-value < 0.032 showing that they are statistically significant. To show statistical significance for game 14, games 1, 2 and 14 are compared. The results of this ANOVA test analysis show a p-value < 0.001, showing that they are statistically significant. A follow up-test, post-hoc test, showed that the statistical significance laid all three games, within games 1, 2 and 14.

4.1.3 Errors

Table 4 – Errors Pair Analyses

Game	1	2	3	4	5	6	7	8	9	10	11	12	13	14
p-value	0.006		0.010	0.036	0.820		0.031	0.029	0.151	0.033	0.047		0.044	
Game Set Combination	(1,2)		(2,3)	(2,4)	(5, 6)		(2,7)	(2,8)	(9,10)	(10,13)	(11,12)		(13, 14)	

For game set A, B, C, E, F and G, this variable averaged lower errors for the changed game than the original game. The only set that was an exception to this is game set D. The following results discussed can be found in Table 4.

To show statistical significance for game 1, games 1 and 2 are compared. The results of this t-test analysis show a p-value < 0.006 showing statistical significance. To show statistical significance for game 3, games 2 and 3 are compared. The results of this t-test analysis show a p-value < 0.01 showing statistical significance. To show statistical significance for game 4, games 2 and 4 are compared. The results of this t-test analysis show a p-value < 0.036 showing statistical significance. To show statistical significance for game 5 and 6, these games are compared. The results of this t-test analysis show a p-value < 0.820 showing that it is not statistically significant. No other combination of games, including the 5th and sixth game, was found to be statistically significant. To show statistical significance for game 7, games 2 and 7 are compared. The results of this t-test analysis show a p-value < 0.031 showing statistical significance. To show statistical significance for game 8, games 2 and 8 are compared. The results of this t-test analysis show a p-value < 0.029 showing statistical significance. To show statistical significance for game 9, games 9 and 10 are compared. The results of this t-test analysis show a p-value < 0.151 showing that these two games are not statistically significant. No other combination of games, including the 9th game, was found to be statistically significant. To show statistical significance for game 10, games 10 and 13 are compared. The results of this t-test analysis show a p-value < 0.033 showing statistical significance. To show statistical significance for game 11 and 12, games 11 and 12 are compared. The results of this t-test analysis show a p-value < 0.047 showing statistical significance. To show statistical significance for game 13 and 14, games 13 and 14 are compared. The results of this t-test analysis show a p-value < 0.044 showing statistical significance.

4.1.4 Cognitive Load

Table 5 – Cognitive Load Pair Analyses

Game	1	2	3	4	5	6	7	8	9	10	11	12	13	14
p-value	0.003	0.450	0.017	0.399	0.575	0.021	0.010	0.011		0.743	0.826		0.910	
Game Set Combination	(1,9)	(1,2)	(3,9)	(3,4)	(5,6)	(6,9)	(7,9)	(8,9)		(10,13)	(11,12)		(13, 14)	

For game set A, B, D and F, this variable averaged a lower cognitive load for the changed game than the original game. The only sets that were an exception to this are game sets C, E and G. The following results discussed can be found in Table 5.

To show statistical significance for game 1, games 1 and 9 are compared. The results of this t-test analysis show a p-value < 0.003 showing statistical significance. To show statistical significance for game 1, games 1 and 2 are compared. The results of this t-test analysis show a $p\text{-value} < 0.450$ showing statistical insignificance. No other combination of games, including the 2nd game, was found to be statistically significant. To show statistical significance for game 3, games 3 and 9 are compared. The results of this t-test analysis show a $p\text{-value} < 0.017$ showing statistical significance. To show statistical significance for game 4, games 3 and 4 are compared. The results of this t-test analysis show a $p\text{-value} < 0.399$ showing statistical significance. No other combination of games, including the 4th game, was not found to be statistically significant. To show statistical significance for game 5, games 5 and 6 are compared. The results of this t-test analysis show a $p\text{-value} < 0.575$ showing statistical significance. No other combination of games, including the 5th, was found not to be statistically significant. To show statistical significance for game 6, games 6 and 9 are compared. The results of this t-test analysis show a $p\text{-value} < 0.021$ showing statistical significance. To show statistical significance for game 7, games 7 and 9 are compared. The results of this t-test analysis show a $p\text{-value} < 0.01$ showing statistical significance. To show statistical significance for game 8 and 9, these games are compared. The results of this t-test analysis show a $p\text{-value} < 0.011$ showing statistical significance. To show statistical significance for game 10, games 10 and 13 are compared. The results of this t-test analysis show a $p\text{-value} < 0.743$ showing statistical significance. No other combination of games, including the 10th game, was found to be statistically insignificant. To show statistical significance for game 11 and 12, these two games are compared. The results of this t-test analysis show a $p\text{-value} < 0.826$ showing statistical insignificance. No other combination of games, including the 11th and 12th game, was found to be statistically significant. To show statistical significance for game 13 and 14, these two games are compared. The results of this t-test analysis show a p-value < 0.910 showing statistical insignificance. No other combination of games, including the 13th and 14th game, was found to be statistically significant.

4.2 NASA TLX Analyses

Table 6 – NASA TLX Results

	MID	END
Mental Demand	40.34	40.31
Physical Demand	4.53	4.25
Temporal Demand	22.69	24.06
Performance	59.66	57.75
Effort	30.84	26.72
Frustration	10.38	12.03
Adjusted Rating	11.24	11.01

The adjusted rating scores were calculated two times: once for the mid-game rating sheets and once for the end game rating sheets and are shown in Table 6. The results of the NASA-TLX Questionnaires showed the same pattern as the quantitative results. The minimum adjusted rating score possible is 0 and the maximum is 100. Therefore, an adjusted rating score of 100 would indicate that this dimension was a very large contributor, imposed a significant weight, to workload; where an adjusted rating score of 0 would indicate that this dimension was not a contributor to workload.

Therefore, given Table 6, we can see that for both the mid games and end of games, performance was the largest mental workload factor imposed on the user, followed by mental demand. The least contributing dimension was physical demand.

The adjusted rating score for the end game is lower than the mid-game, showing the same trend as the CL results that by the end of the games, CL lowered from ~0.5 to ~0.4. This trend also follows for each factor rated. The only exception to this is Temporal Demand and Frustration. This would imply that despite an average game taking no longer than 70 seconds, even though the participants had up to 150 seconds, the time stress was significant enough, and the game increased in difficulty enough to increase temporal demand, as well as frustration. Even though the cognitive load was no different, this shows that other factors were at play.

4.3 Questionnaire Analyses

A third of the participants used tracing strategies to remember the sequence and shapes. Another third used a speak aloud, whether aloud verbally or in their head. The remaining third used a variety of different strategies, including counterclockwise and clockwise patterns, associating locations, up down left right, associating numbers instead of shapes and some coding strategies, like the one Miller described [20]. 6 participants found the first half of the experiment was more difficult as they had to get accustomed with the game and used to coping with changes and sequences; easier as they played more games. Nine participants found no change in difficulty. The remaining found the second half more difficult for many reasons, including mental fatigue, losing focus, compounded errors and compounded changes. The most prominent, most challenging change for most participants to cope with was the square, spread out or when these changes were combined because of the extra distance to move their thumb to press the button. This is because of the change in their tracing or coding strategy and, therefore, the change in their mental model. There was also stated difficulty in the stimulus button being situated on the right then the left due to right-handedness where the stimulus on the other side would have improved performance since they would have been able to use both hands.

Chapter 5 – Discussion

In this chapter, the results of the experiment are discussed. Specifically, results are considered in relation to the hypothesis questions posed earlier. As well as, in comparison and reference to some literature covered previously.

5.1 Hypotheses

RQ1: Does following the Gestalt law of Figure-Ground, Proximity or Closure increase cognitive load, speed, and performance?

In terms of time, the trials that introduced a change (game 2,4,8,10,12 and 14) are lower than their unchanged counterparts, except for one game, game 6. Given this, we can deduce that following the Gestalt Principles of Figure-Ground, Closure and Proximity decreases speed. In terms of errors, the trials that introduced a change (game 2,4,6,10,12 and 14) experienced fewer errors than their unchanged counterparts, except for one game, game 8. Given this, we can deduce that following the Gestalt Principles of Figure-Ground, Closure, and Proximity improve performance. In terms of stimulus reaction time, 3 of the changed games are lower than their unchanged counterparts, being games 2,4 and 12. However, changes in games (6, 8, 10, and 14) were slightly greater than their unchanged game counterparts. Given this, we can deduce that following the Gestalt Principle of Proximity lowers cognitive load because games 4 and 12 each tested the principle of Proximity. These can be seen in table 2.

RQ2: Does compounding different changes/principles affect cognitive load, speed, and performance?

As previously explained, given that the changed games were lower than their unchanged counterparts for time and errors, as can be seen in table 2, compounding changes do not affect speed or performance. Although, given that some changed-unchanged game pairs, one of them being the compounded changes, were higher than their counterparts in terms of stimulus reaction time, we can deduce that compounded changes do increase cognitive load.

RQ3: Does changing the placeholders (change for class 4 in games 8 and 14) affect a user's mental model, given that they remembered the original placeholders by their position and not their label?

Given the exceptions of errors for game 8, as can be seen in table 2, it suggests that a memory change does increase the number of errors as user's are readjusting their muscle memory patterns.

RQ4: Do participants encode the sequence differently than they are presented? Does encoding in this fashion affect performance, speed and cognitive load?

20 percent of participants mentioned that to remember the sequence presented to them, they encoded the sequence in these ways:

- If the sequence started with XO, I ignored those shapes and focused on the last 1-5 shapes
- I also said “oh” in my head instead of “circle” to save a syllable when reading the “O” shape.
- memorizing of similar sub sequence
- noticing patterns
- Memorizing every 2 shapes as one finger movement.
- Memorizing the finger swipes instead of the individual clicks.
- Square = D, X = X, Triangle = T, Circle = O || Circle O, Square S, Triangle T, X X.
- Previous sequence + new shape

RQ5: Does muscle memory come into play when trying to remember a sequence of steps to complete a task?

At least a third of participants memorized the sequences by tracing the increasing pattern with their input finger or thumb as well as eleven participants mentioned that the square change was the most difficult to cope with. This would indicate that muscle memory came into play because they could no longer rely on their tracing strategy and, therefore, tracing memory to remember the sequence at hand.

As previously discussed, in a study by Haji [24], it was found that their performance scores (and intrinsic cognitive load decreased) increased as participants progress in the study. It was found to be the case because they began to establish mental models that reduced their intrinsic cognitive load. Given that the performance scores did not vary significantly between the first to last game, it is possible that the task was not demanding enough to warrant creating a mental model versus the task that was presented in the study by Haji [24]. Although, we can confirm this trend given the NASA-TLX scores that the end game ratings were lower than the mid game ratings.

Secondly, we can see that as performance increased (time decreased), cognitive load also decreased, as suggested by the second method of utilizing the dual task paradigm as suggested by Brunken [9].

Thirdly, similarly to the studies conducted by [30] [6] [39] and [17] seen that results flattened out after multiple trials when encountered with a different interface or layout. These results were not observed in the changed games because new variations were introduced for each. Although it is observed in the unchanged games (1, 3, 5, 7, 9, 11, and 13), as these were the same interfaces repeated. By the 13th game, the average time scores were reduced from ~70 to ~60, errors were reduced from ~2 to ~1.2 and cognitive load from ~0.5 to ~0.4. This can be seen in Figure 24.

Fourthly, given that the variation in the performance scores was not very different from the beginning of the experiment to the end, it can be assumed that the participants were able to cope with the additional changes of the interface, as well as the transition from a novice to an expert player, given that it was presented in a simple to complex strategy [56], beginning with low element interactivity (isolated elements) moving into high element interactivity, implementing multiple changes.

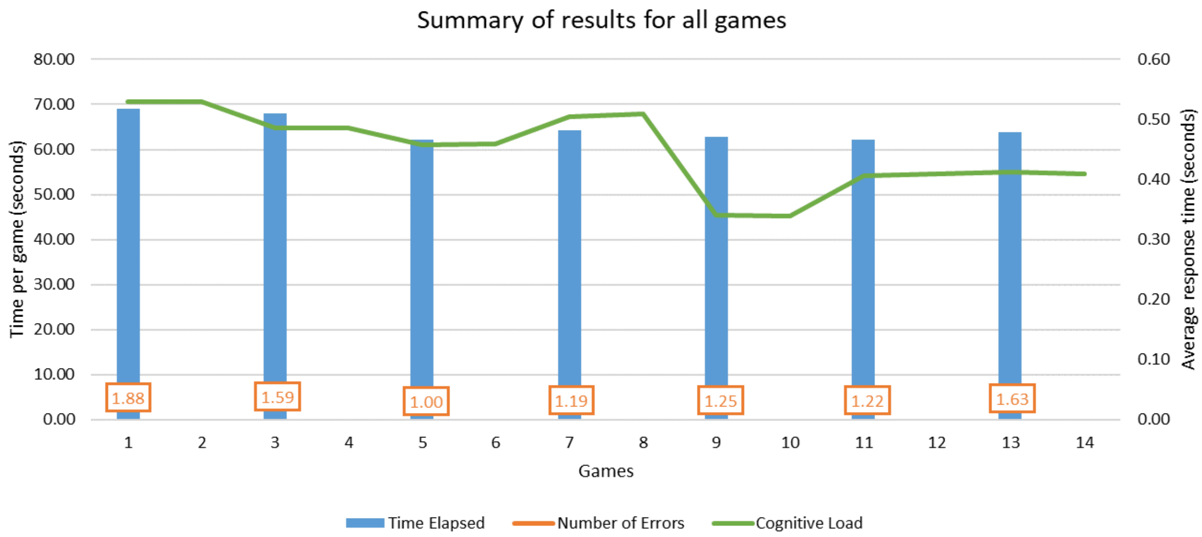


Figure 24 – Steady Results in Unchanged Games

Lastly, as a result, this research study demonstrated that specific Gestalt principles are chosen, and the degree to which they were manipulated on the user interface was profound enough to elicit a slight but not significant enough change in performance in participants.

To conclude, given that statistical significance could not be found directly and had to be inferred from combinations of games, we must infer that the data is trending in a sense to support the conclusions I have made. To suppose these conclusions with 100% certainty, the experiment would need to be reconducted anew while gathering additional participants to increase the statistical power and augment the chance of finding a significant effect [18].

5.2 Limitations

Given the COVID-19 global pandemic, there were limitations in testing in-person that caused an inability to use the EEG for an objective cognitive load collection tool. To circumvent this, the NASA-TLX rating method in combination with the dual-task paradigm are used as an alternative to deduce cognitive load.

Chapter 6 – Conclusion

In conclusion, it was found that the seven games in which the inputs were changed did not elicit a large change in performance or cognitive load, other than a few exceptions. A series of ANOVA and t-tests were conducted to evaluate the significance of these pairs of games. From this, it was found that only certain pairs were significant. From this, it can be concluded that the user interface modifications tested based on Gestalt Principle of Figure-Ground, Proximity and Closure did not elicit a significant effect from the user. Therefore, changes similar to those tested could be implemented on an interface and given to users with little concern for negative repercussions.

Although, given the statistical analyses tests, to state this with 100% certainty, the experiment would need to be reconducted with additional participants to reevaluate the statistical significance of the data collected.

It is acknowledged that using the dual-task paradigm for collecting cognitive load data collection is less than ideal. However, given circumstances surrounding COVID-19, the EEG was not a viable measure to collect cognitive load data.

6.1 Future Work

As mentioned previously, a beginning step to future work would include the reconduction of the above-described experiment with additional participants. This would be to ensure that the statistical power is high enough which would in turn elicit a higher chance of finding statistical significance in the data [18]. Additionally, the experiment should be reconducted utilizing an EEG to collect cognitive load data once this option is viable. Along with the inclusion of a larger variability of age groups to ensure no age group is excluded.

Moreover, the following logical steps to this experiment would be to test other, larger changes in the same fashion, to further understand how various changes affect the user. As well as, to establish a guideline to indicate which changes can be made with or without affecting the user. Conversely, to what extent one or multiple changes can effect a user.

As can be seen from these variations of skill acquisition stages, many researchers have built upon and modified these stages for various tasks, goals, and disciplines. Although, it would be of interest to create a system in which it would have the capacity to measure at what stage of proficiency the user is. As well as automatically feeding this information into an adaptive user interface to apply design guidelines, discussed above, to adapt them to the user on the fly. This adaptive user interface could utilize Bayesian Inference to automatically account for previous user experiences [3]. This adaptive user interface could also act as an instructor figure suggested by Dreyfus [16] to point these elements out for the learner.

References

1. Alpert, F. F., et al. "533 Evaluation of subjective scales for measuring mental workload: literature review." (2018): A613-A613.

2. Apple https://www.apple.com/ca/iphone/

3. Barrette, Rachelle, and Ratvinder Grewal. "Adaptive User Interfaces and the Use of Inference Methods." Computational Science and Techniques 8 (2021): 616-624.

4. Bell https://www.bell.ca/

5. Benner, P. (1982). From novice to expert. *American Journal of nursing*, *82*(3), 402-407.

6. Bi, X., & Zhai, S. (2016, May). IJqwerty: What difference does one key change make? Gesture typing keyboard optimization bounded by one key position change from Qwerty. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 49-58).

7. Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated–interactive elements effect and levels of learner expertise: Experimental evidence from an accountancy class. *Instructional Science*, *38*(3), 277-287.

8. Blayney, P., Kalyuga, S., & Sweller, J. (2015). Using cognitive load theory to tailor instruction to levels of accounting students' expertise. *Journal of Educational Technology & Society*, *18*(4), 199-210.

9. Brunken, Roland, Jan L. Plass, and Detlev Leutner. "Direct measurement of cognitive load in multimedia learning." Educational psychologist 38.1 (2003): 53-61.

10. BWD https://bestwebdesign.co.za/design-principles-gestalt-psychology/

11. Cacitti, L., & Besnard, D. (2003). Interface changes generating accidents. A schema-driven approach of negative transfer. *TECHNICAL REPORT SERIES-UNIVERSITY OF NEWCASTLE UPON TYNE COMPUTING SCIENCE*.

12. Chang, D., & Nesbitt, K. V. (2006, April). Developing Gestalt-based design guidelines for multi-sensory displays. In *Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop-Volume 57* (pp. 9-16). Australian Computer Society, Inc.

13. Chen, Daniel, and Roel Vertegaal. "Using mental load for managing interruptions in physiologically attentive user interfaces." CHI'04 extended abstracts on Human factors in computing systems. 2004.

14. Cockburn, A., Gutwin, C., Scarr, J., & Malacria, S. (2014). Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys (CSUR)*, *47*(2), 1-36.

15. DIS, I. (2009). 9241-210: 2010. Ergonomics of human-system interaction-Part, 210.

16. Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition* (No. ORC-80-2). California Univ Berkeley Operations Research Center.

17. Dunlop, M., & Levine, J. (2012, May). Multidimensional pareto optimization of touchscreen keyboards for speed, familiarity and improved spell checking.

In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2669-2678).

18. Field, Andy. Discovering statistics using IBM SPSS statistics 5th edition. Sage edge, 2017.

19. Fleischmann, Marvin, et al. "The role of software updates in information systems continuance—An experimental study from a user perspective." Decision Support Systems 83 (2016): 83-96.

20. Fuchs, Sven. "HCI 450–Foundations of Usability Research Review How magical is Miller's famous Seven for HCI issues?."

21. Garett, Renee, et al. "A literature review: website design and user engagement." Online journal of communication and media technologies 6.3 (2016): 1.

22. Google https://store.google.com/ca/category/phones

23. Graham, L. (2008). Gestalt theory in interactive media design. *Journal of Humanities & Social Sciences*, *2*(1).

24. Haji, F. A., Rojas, D., Childs, R., de Ribaupierre, S., & Dubrowski, A. (2015). Measuring cognitive load: performance, mental effort and simulation task complexity. *Medical education*, *49*(8), 815-827.

25. Hart, S. G. (1986). NASA Task load Index (TLX). Volume 1.0; Paper and pencil package.

26. Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner Jr, A. C., Zaklade, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human factors*, *34*(4), 429-439.

27. IBM https://en.wikipedia.org/wiki/List_of_mergers_and_acquisitions_by_IBM

28. In-Car User Interfaces https://www.nngroup.com/articles/why-consumer-products-have-bad-ux/

29. Johnson, J. *Designing with the mind in mind: simple guide to understanding user interface design guidelines*. Morgan Kaufmann.

30. Jokinen, J. P., Sarcar, S., Oulasvirta, A., Silpasuwanchai, C., Wang, Z., & Ren, X. (2017, May). Modelling learning of new keyboard layouts. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 4203-4215).

31. Koppel, R., Metlay. J.P., Cohen, A., Abaluck, B., Localio, A.R., Kimmel, S.E., and Strom B.L. (2005): "Role of Computerized Physician Order Entry Systems in Facilitating Medication Errors," *Journal of the American Medical Association* Vol. **293** No. 10 (March 9), 1197-1203.

32. Leppink, Jimmie, and Angelique van den Heuvel. "The evolution of cognitive load theory and its application to medical education." Perspectives on medical education 4.3 (2015): 119-127.

33. Ling, J., & Van Schaik, P. (2002). The effect of text and background colour on visual search of Web pages. *Displays*, *23*(5), 223-230.

34. Ling, J., & Van Schaik, P. (2006). The influence of font type and line length on visual search and information retrieval in web pages. *International journal of human-computer studies*, *64*(5), 395-404.

35. Ling, J., & van Schaik, P. (2007). The influence of line spacing and text alignment on visual search of web pages. *Displays*, *28*(2), 60-67.

36. Mark, G., Gudith, D., & Klocke, U. (2008, April). The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 107-110).

37. McIlroy, Stuart, Nasir Ali, and Ahmed E. Hassan. "Fresh apps: an empirical study of frequently-updated mobile apps in the Google play store." Empirical Software Engineering 21.3 (2016): 1346-1370.

38. Möller, B., Brezing, C., & Unz, D. (2012). What should a corporate website look like? The influence of Gestalt principles and visualisation in website design on the degree of acceptance and recommendation. *Behaviour & Information Technology*, *31*(7), 739-751.

39. Niemelä, M., & Saariluoma, P. (2003). Layout attributes and recall. *Behaviour & information technology*, *22*(5), 353-363.

40. Norman, D. (2013). The design of everyday things: Revised and expanded edition. Basic books.

41. Paas, F. G., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and motor skills*, *79*(1), 419-430.

42. Paas, Fred, Alexander Renkl, and John Sweller. "Cognitive load theory and instructional design: Recent developments." Educational psychologist 38.1 (2003)

43. Plouznikoff, Nicolas, Alexandre Plouznikoff, and J-M. Robert. "Artificial grapheme-color synesthesia for wearable task support." *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*. IEEE, 2005.

44. Preece, J., Rogers, Y., & Sharp, H. (2015). Interaction design: beyond human-computer interaction. John Wiley & Sons.

45. Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational psychologist*, *38*(1), 15-22.

46. Renkl, A., Atkinson, R. K., & Maier, U. H. (2000). From studying examples to solving problems: Fading worked-out solution steps helps learning. In *Proceeding of the 22nd Annual Conference of the Cognitive Science Society* (pp. 393-398).

47. Ripalda, D., Guevara, C., & Garrido, A. (2020, July). Framework Based on Gestalt Principles to Design Mobile Interfaces for a Better User Experience. In *International Conference on Applied Human Factors and Ergonomics* (pp. 158-165). Springer, Cham.

48. Rogers https://www.rogers.com/

49. Roto, V. E. (2011). User experience white paper. http://www. allaboutux. org/uxwhitepaper.

50. Samsung https://www.samsung.com/ca/mobile/

51. Seel, N. M. (1995). Mental Models, Knowledge Transfer and Teaching Strategies. *Journal of Structural Learning, 12(3), 197-213*

52. Silva, Carlos, et al. "A use error taxonomy for improving human-machine interface design in medical devices." ACM SIGBED Review 16.2 (2019): 24-30.

53. Simon https://www.memozor.com/simon-games/simon-game

54. Toptal https://www.toptal.com/designers/ui/gestalt-principles-of-design

55. User Testing https://www.usertesting.com/blog/gestalt-principles#common

56. Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical education*, *44*(1), 85-93.

57. Wierwille, W. W., & Casali, J. G. (1983, October). A validated rating scale for global mental workload measurement applications. In *Proceedings of the human factors society annual meeting* (Vol. 27, No. 2, pp. 129-133). Sage CA: Los Angeles, CA: Sage Publications.

Appendix A – Pre-Questionnaire

LaurentianUniversity
UniversitéLaurentienne

Demographics Questionnaire

1. Age

- ◯ 18-25
- ◯ 26-39
- ◯ 40+
- ◯ Prefer not to answer

2. Program of study or Department ____________________

3. How many different devices do you use per day? (Select One)

	0	1-2	3-4	5+
Number of Devices Used/Day	◯	◯	◯	◯

4. What types of devices do you typically use per day? (Select Many)

- ☐ Cellphone
- ☐ Laptop/Desktop
- ☐ Tablet
- ☐ Wearables (e.g Smartwatch)
- ☐ Game Consoles (Any)
- ☐ Other ____________________

5. How proficient do you consider yourself with these devices?
1(Not Very) - 5 (Very)

	NA	1	2	3	4	5
Cellphone	◯	◯	◯	◯	◯	◯
Laptop/Desktop	◯	◯	◯	◯	◯	◯
Tablets	◯	◯	◯	◯	◯	◯
Wearable	◯	◯	◯	◯	◯	◯
Game Console	◯	◯	◯	◯	◯	◯
Other	◯	◯	◯	◯	◯	◯

6. How long do you use each device per day?

	NA	<1 hr	1-3 hrs	4-5 hrs	6-7 hrs	8+ hrs
Cellphone	◯	◯	◯	◯	◯	◯
Laptop/Desktop	◯	◯	◯	◯	◯	◯
Tablet	◯	◯	◯	◯	◯	◯
Wearable	◯	◯	◯	◯	◯	◯
Game Console (Any)	◯	◯	◯	◯	◯	◯
Other	◯	◯	◯	◯	◯	◯

7.How proficient do you consider yourself to be with the following tasks?

1 (Not Very) - 5 (Very)

	NA	1	2	3	4	5
Input using a keyboard	◯	◯	◯	◯	◯	◯
Input using touchscreen	◯	◯	◯	◯	◯	◯
Searching the web	◯	◯	◯	◯	◯	◯

8. Do you prefer?

◯ Touchscreen
◯ Mouse/keyboard/trackpad
◯ Controllers (Any)
◯ No Preference

Appendix B – Post-Questionnaire

Laurentian University
Université Laurentienne

Post Questionnaire

1. What strategy did you use in remembering the sequences, if any?

2. What strategy did you use in remembering the shapes and locations of the inputs, if any?

3. Did you find an overall change in difficulty throughout the experiment?

4. Were there a (or multiple) visual changes that were most prominent and challenging for you?

5. Did you use a mobile device, tablet or tablet computer during this study?

- ◯ Mobile Device (Cell Phone)
- ◯ Tablet
- ◯ Laptop
- ◯ Desktop
- ◯ Other

Appendix C – NASA TLX Rating & Weighting Forms

LaurentianUniversity
UniversitéLaurentienne

NASA TLX

Participant ID [ ] Task [ ] Date [ ]

Very Low … Very High

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Mental Demand	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Physical Demand	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Temporal Demand	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Performance	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Effort	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Frustration	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

LEGEND

Mental Demand: How Mentally demanding was the task?
Physical Demand: How Physically demanding was the task?
Temporal Demand: How hurried or rushed was the pace of the task?
Performance: How successful were you in accomplishing what you were asked to do?
Effort: How hard did you have to work to accomplish your level of performance?
Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Cards

For use by researcher only

25	○ Frustration ○ Effort	33	○ Temporal Demand ○ Frustration
26	○ Performance ○ Mental Demand	34	○ Temporal Demand ○ Effort
27	○ Performance ○ Temporal Demand	35	○ Physical Demand ○ Frustration
28	○ Mental Demand ○ Effort	36	○ Performance ○ Frustration
29	○ Mental Demand ○ Physical Demand	37	○ Physical Demand ○ Temporal Demand
30	○ Effort ○ Physical Demand	38	○ Physical Demand ○ Performance
31	○ Frustration ○ Mental Demand	39	○ Temporal Demand ○ Mental Demand
32	○ Effort ○ Performance		

Tally for each factor

Effort		Performance	
Mental Demand		Physical Demand	
Temporal Demand		Frustration	

Appendix D – Consent Form

LaurentianUniversity
UniversitéLaurentienne

Participant Consent Form

This document obtains an individual's consent to participate in a research study conducted by Rachelle Barrette under the supervision of Dr. Ratvinder Grewal from the Department of Mathematics and Computer Science at Laurentian University.

RESEARCH STUDY PURPOSE

This research study focuses on how users tolerate a change in a user interface. How and to what degree does altering the user interface of a device have on its users, their performance, and the outcomes of their interactions?

RESEARCH STUDY PROCEDURES

Before beginning the experiment, a demographics questionnaire will be administered that will collect data about the participants' technology usage and background experiences. Participants will then complete a series of trials on a common user interface. In each trial, the participant will see a sequence of buttons to enter at the top of the interface; the participant will then be asked to enter in the sequence, as shown, using the buttons provided on the user interface. These buttons will be manipulated graphically in terms of size, line thickness, spacing, location and other visual variables by the researcher. Concurrently to this, there will also be a dual task to measure cognitive load, where the participant will have to respond to a secondary stimulus presented on the screen by clicking the button when it appears on the user interface, as quickly as possible. Afterwards, a post questionnaire will be completed by the participant. This experiment will be conducted on the web on (mm/dd/yyyy): (/   /   ) and will have a duration of approximately thirty (30) minutes. Rachelle Barrette will be the sole observer and research experiment conductor for this research experiment.

COLLECTION OF EMAIL ADDRESS

Upon completion of the research project the results generated will be available to the participants concerning the effects of interacting with differing user interface designs with the same functionality. A one-page results summary will be sent to the participants via the email they provided at the start of the experiment. This is not required; the participant can choose to omit to share their email. The participant will not be penalized for this.

POTENTIAL RESEARCH STUDY RISKS

Participants will not be exposed to any medical substances or hazardous equipment. There are possible risks to privacy; multiple measures are in place to prevent this. Firstly, no personally identifiable information is collected about the participant apart from their name and email address. The forms that do collect this information will be password protected and stored safely on the LU GDrive and only accessible by the researcher.

POTENTIAL BENEFITS TO PARTICIPANTS

Benefits to the participants may include increased interest and awareness of interface design and how users interact with interfaces.

CONFIDENTIALITY & DATA STORAGE

All efforts will be made to ensure the confidentiality of the participant as well as any identifying information of the participant that is obtained in connection with the study. Information obtained during this research study will be stored in a database. The connection to this database will be secured by an SSL connection with a signed certificate. The results in this database will be encrypted using 256-bit AES encryption technology. Questionnaires, consent forms and email addresses, will be stored in a private folder on Laurentian University's Google Drive, will be password protected and will only be accessible by the researcher. The data will be destroyed after 5 years of the date that the experiment was conducted.

PARTICIPATION AND WITHDRAWAL

At any time throughout the research study process the participant can withdraw their consent to participate in this research study without fear of consequence.

ISSUES/ETHICAL CONCERNS

Issues or ethical concerns held by the participant regarding this research study can be addressed to the Ethics Department at Laurentian University:

Research Ethics Officer
Office of Research Services

Telephone: 705-675-1151 ext 3681 or 2436

Toll Free: 1-800-461-4030

E-mail: ethics@laurentian.ca

Any questions and/or concerns regarding this research study can be directed to:
Rachelle Barrette: rbarrette1@laurentian.ca;
Dr. Ratvinder Grewal: rgrewal@cs.laurentian.ca; Phone No. 1-705-675-1151 ext: 2351

SIGNATURE OF RESEARCH PARTICIPANT

I have read the information provided for the study "Tolerance of User Interface Design Alterations" as described herein. My questions have been answered to my satisfaction and I agree to participate in this study. I have been given a copy of this form.

Name of Participant (Please Print or Type)	
Signature of Participant	
Email Address	
Date(mm/dd/yyyy)	

Appendix E – Script

Good Morning/Afternoon/Evening,

I'd like to begin by sharing my deepest gratitude with you for your participation in my experiment. If it wasn't for you, the participant, this would not be possible. So, thank you very much for taking time out of your day to participate, especially in these uncertain times, with COVID-19 and the beginning of this unique new semester.

To introduce myself, my name is Rachelle; I am a Graduate Student in the Master of Science in Computational Science program at Laurentian University. Just to reiterate from my email, I will have my microphone and camera on at all times but this is not required of you, although I do ask for you to have your microphone to facilitate easier communication.

The experiment I have designed can be compared to a Simon game. In this game, a series colours would appear in an incrementing series (1 colour, two colours, three colours, etc.), and the player would be required to remember and enter those colours in the sequence they appeared on the gamepad, if they didn't, then they would be required to restart. The experiment I have designed uses shapes instead of colours and is restricted to 7 items in the sequence instead of an unlimited number. And in my experiment, I ask of you to play a total of 14 games and to fill out 4 questionnaires.

This is what the interface will look like. Firstly, across the top, you have the timer bar, you have 150 seconds in total to complete the sequence, once this bar runs out, you are out of time and will have to restart that sequence. The Pause button; if you need to pause for whatever reason, click this. You'll notice on the right hand side a white box; it will appear on the interface every so often, as soon as this appears, click it, even if you are in the midst of entering a sequence. And on the left you have the input buttons, once you see the sequence that appears under the timer bar, enter it in using these inputs. If you make an error, you will have to repeat that sequence; the interface will indicate to you the first shape that you entered incorrectly then will proceed to allow you to re-enter that sequence. The underlines indicate how many inputs are required; in case you forget how many inputs the application is expecting.

Before we begin, I am required to ask for your consent in participating in my experiment. There are no risks to your physical health or safety but there are risks to your privacy. I have multiple measures in place to ensure your privacy is kept. The only personally identifiable information that I am collecting is your email. This information will be stored, password protected using a password that only I know on Laurentian University's secure Google Drive. That being said, you do have the right to not share any of the information that is asked if you do not feel comfortable, including your email, age, program of study/department, etc… I ask for your email to share a summary of your results after you are done the experiment; although, again, this is not required, you have a right to refuse to share this information. You also have a right to withdraw your consent to this experiment and terminate the study at any point. If you have any questions, please feel free to ask them now. If you consent, I will share with you a folder on Google Drive that contains all of the questionnaires required for this experiment. Please download the consent form, fill it out using a PDF viewer and "print" it as a pdf and reupload it to the same google drive folder. If you do not, I thank you for your interest and wish you the best of luck for the upcoming semester. I will also ask you to fill out a first questionnaire that will ask some basic questions about your technology usage, proficiency level. Please download it, fill it out, and reupload it to google drive.

Alright, before we start the experiment, I'll let you play around on a test game. Due to some programming limitations, there are a few things we need to change on your device before conducting the experiment. Once you have the site open, leave your device upright/portrait, request the desktop site for your device. Once you do this, the site should reload and will look like a blue box on a white screen. Click the full screen button at the bottom right. Now turn off the rotation lock for your device, and turn your device to the side/landscape mode.

Before we begin the experiment, after the 8th trial, we will take a break to fill out a NASA TLX questionnaire. At that point, a big yellow screen will appear, DO NOT CLICK this button; it automatically unpauses the game. Just switch to zoom and we will go ahead with the survey.

**** AT MID POINT**** I will ask you to download and fill out the first rating scale and reupload it to the google drive folder. In this survey, rate each term from one to 21 as the strength of these throughout these first trials. Now go back to the game and click this yellow button.

To conclude the experiment, I will begin by showing you a series of two terms on the screen and you tell me which was most evident to you during the experiment. Secondly, I will ask you to download and fill out the second rating scale and reupload it to the google drive folder. Lastly, there is one more questionnaire, a post-questionnaire, to fill out, please fill it out and reupload it to the google drive folder as well.

Alright and that concludes the experiment. Once again, thank you for your participation, it is extremely appreciated. Unless you have any questions; I will let you go at this point in time. Have a great day/night/evening and best of luck in the upcoming semester.

Appendix F - Recruitment Poster

Department of *Mathematics and Computer Science*
Laurentian University

LaurentianUniversity
UniversitéLaurentienne

PARTICIPANTS NEEDED FOR RESEARCH IN

Tolerance of User Interface Design Alterations

We are looking for volunteers to take part in a study investigating the tolerance of user interface design alterations.

As a participant in this study, you would be asked to play a series of games similar to the "*Simon*" Game on a user interface hosted on the web and fill out multiple questionnaires.

Your participation would involve *1 session*,
of which is approximately *(30)* minutes.

For more information about this study, or to volunteer for this study, please contact:

Rachelle Barrette
Department of Mathematics and Computer Science
Email: rbarrette1@laurentian.ca

via

Zoom (Remote Delivery)

This study has been reviewed by and received ethics clearance through, Laurentian University's Research Ethics Committee.

Appendix G – Ethics Approval

LaurentianUniversity
UniversitéLaurentienne

APPROVAL FOR CONDUCTING RESEARCH INVOLVING HUMAN SUBJECTS
Research Ethics Board – Laurentian University

This letter confirms that the research project identified below has successfully passed the ethics review by the Laurentian University Research Ethics Board (REB). Your ethics approval date, other milestone dates, and any special conditions for your project are indicated below.

TYPE OF APPROVAL / New X / Modifications to project / Time extension	
Name of Principal Investigator and school/department	Rachel Barrette, Ratwinder Grewal, and Amy Doan, supervisers, Math and Computer Sciences
Title of Project	User Tolerance in User Interfaces (Remote Delivery)
REB file number	6020713
Date of original approval of project	September 1st, 2020
Date of approval of project modifications or extension *(if applicable)*	
Final/Interim report due on: *(You may request an extension)*	September 1st, 2021
Conditions placed on project	

During the course of your research, no deviations from, or changes to, the protocol, recruitment or consent forms may be initiated without prior written approval from the REB. If you wish to modify your research project, please refer to the Research Ethics website to complete the appropriate REB form.

All projects must submit a report to REB at least once per year. If involvement with human participants continues for longer than one year (e.g. you have not completed the objectives of the study and have not yet terminated contact with the participants, except for feedback of final results to participants), you must request an extension using the appropriate LU REB form. In all cases, please ensure that your research complies with Tri-Council Policy Statement (TCPS). Also please quote your REB file number on all future correspondence with the REB office.

Congratulations and best wishes in conducting your research.

Rosanna Langer

Rosanna Langer, PHD, Chair, *Laurentian University Research Ethics Board*

Appendix H – Steps of a Game

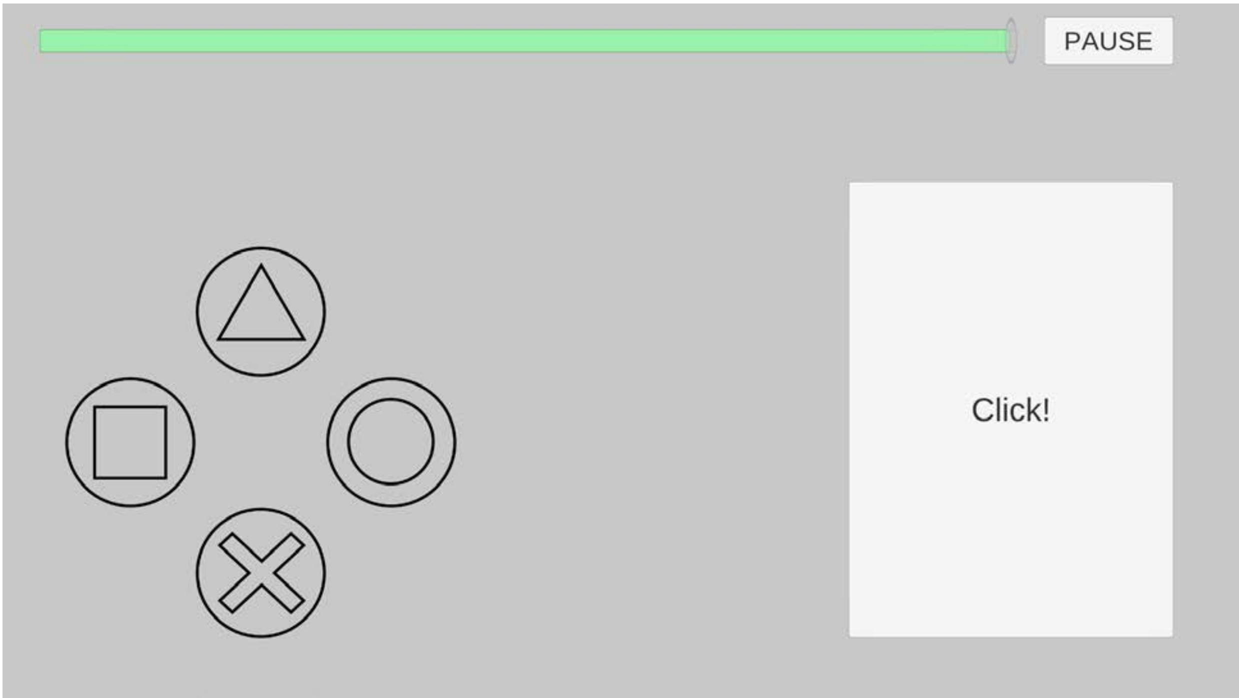


Figure 25 – Before the Game Starts (No Sequence Presented Yet)

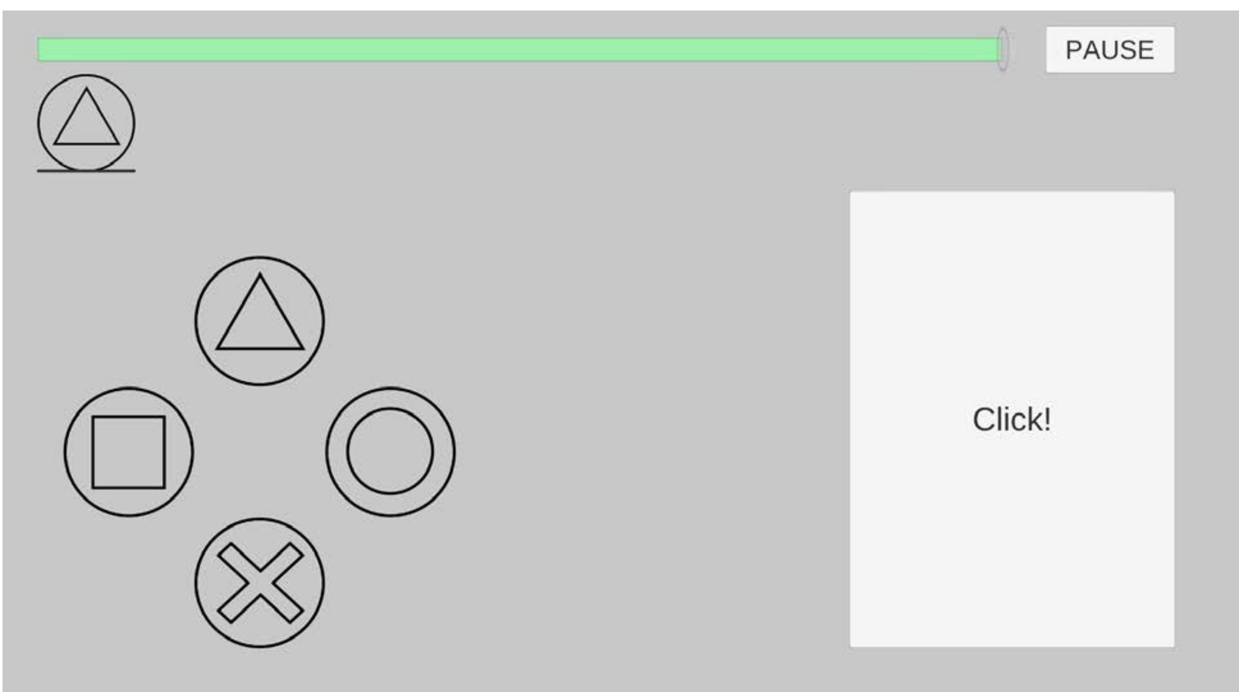


Figure 26 – First Sequence Appears, 1 Shape/Input to Enter

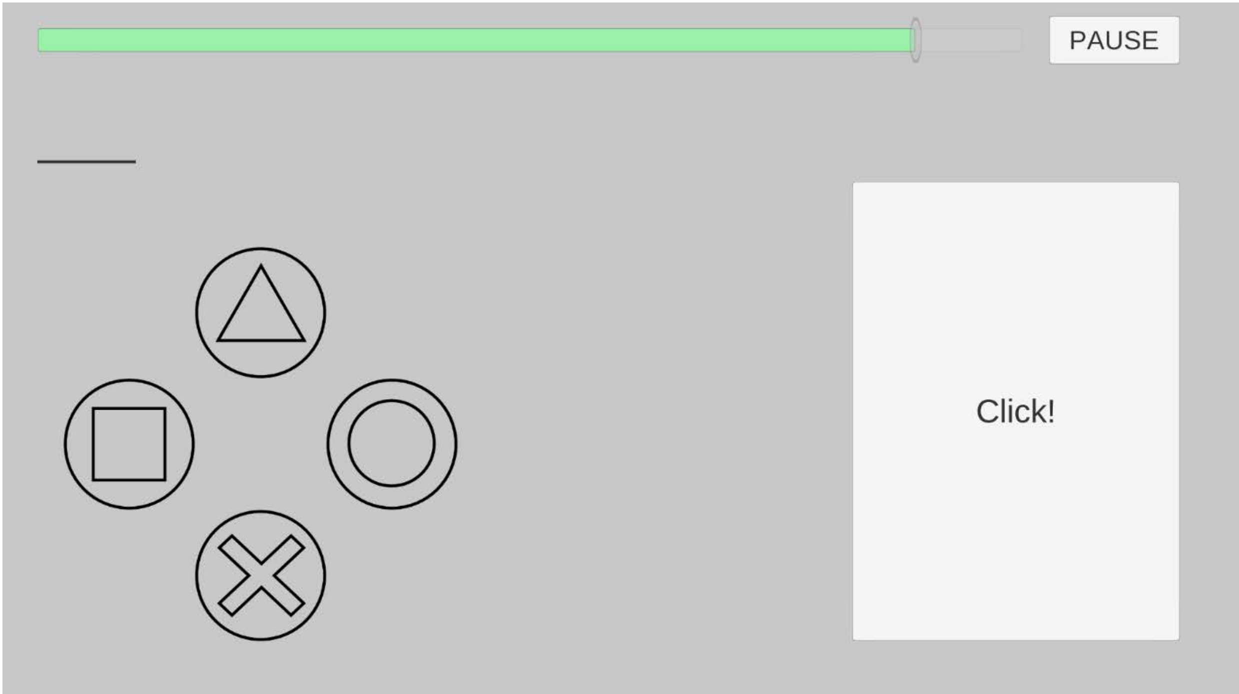


Figure 27 – Awaiting Input of One Shape

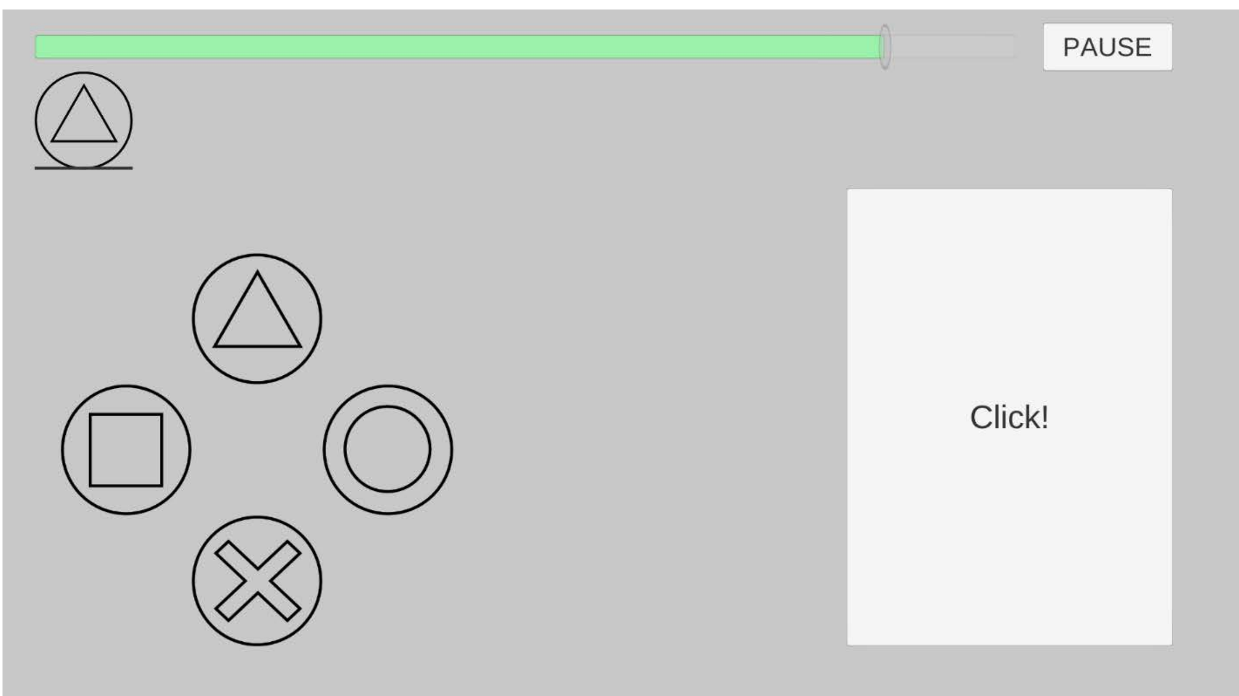


Figure 28 – Successful Input of One Shape

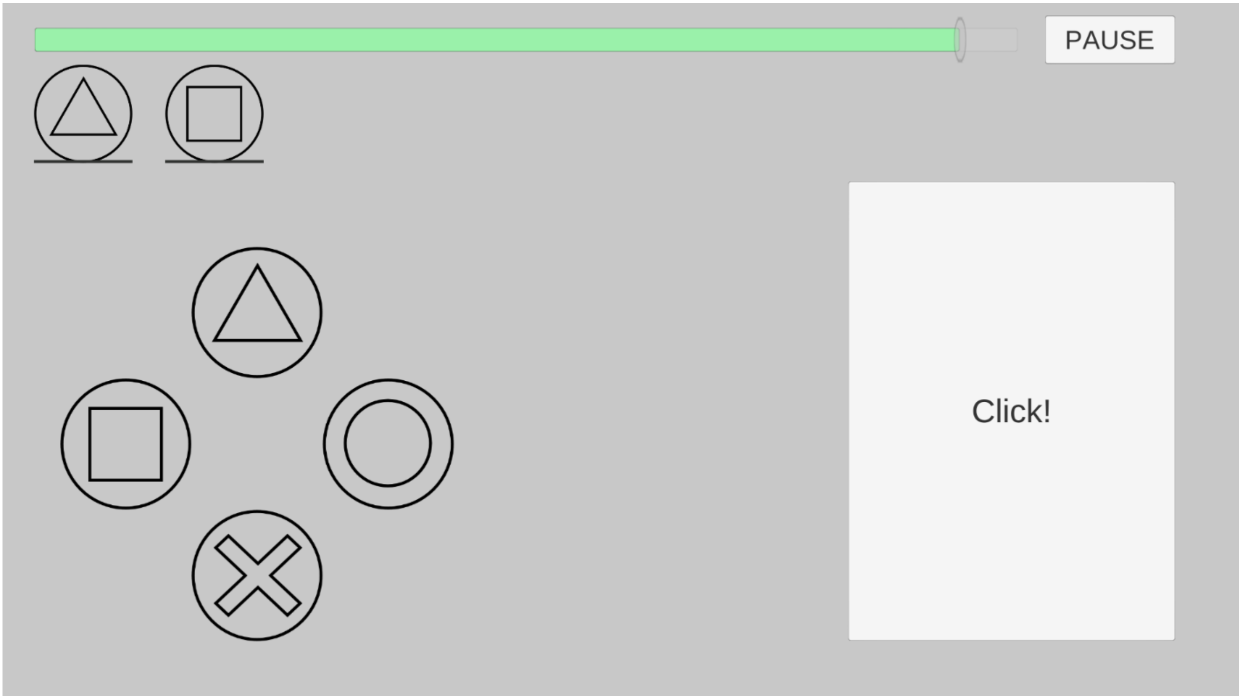


Figure 29 – First Sequence Appears, 2 Shapes/Inputs to Enter

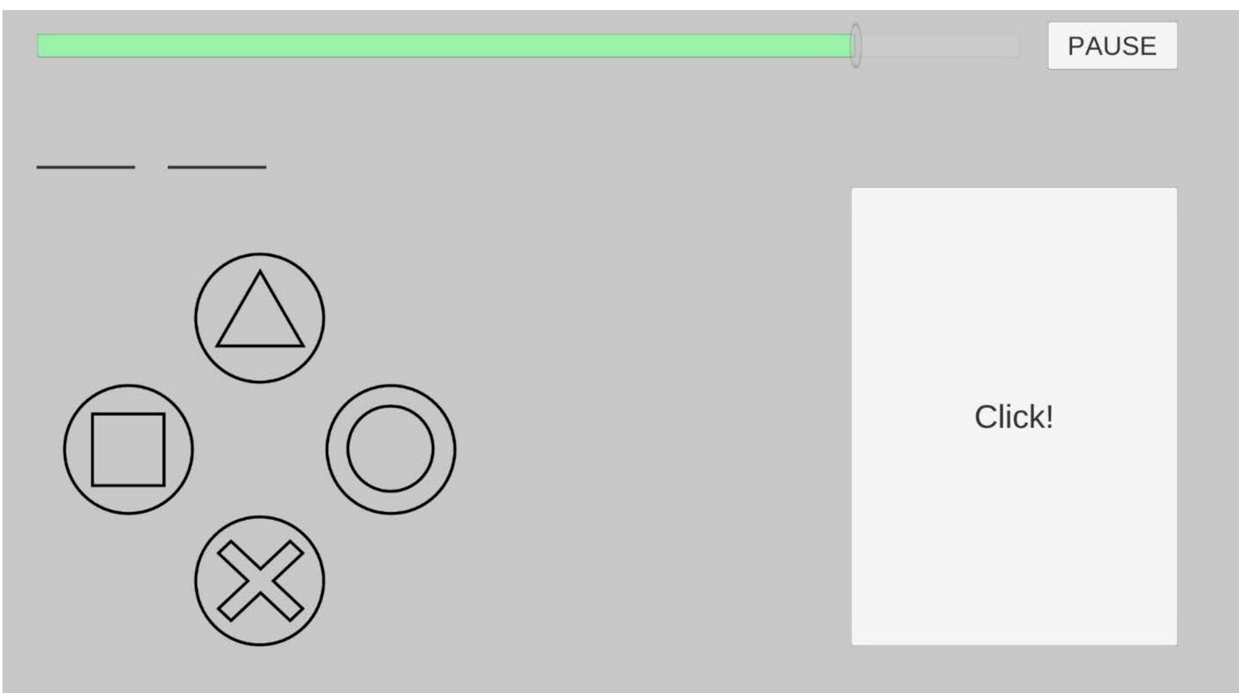


Figure 30 – Awaiting Two Inputs

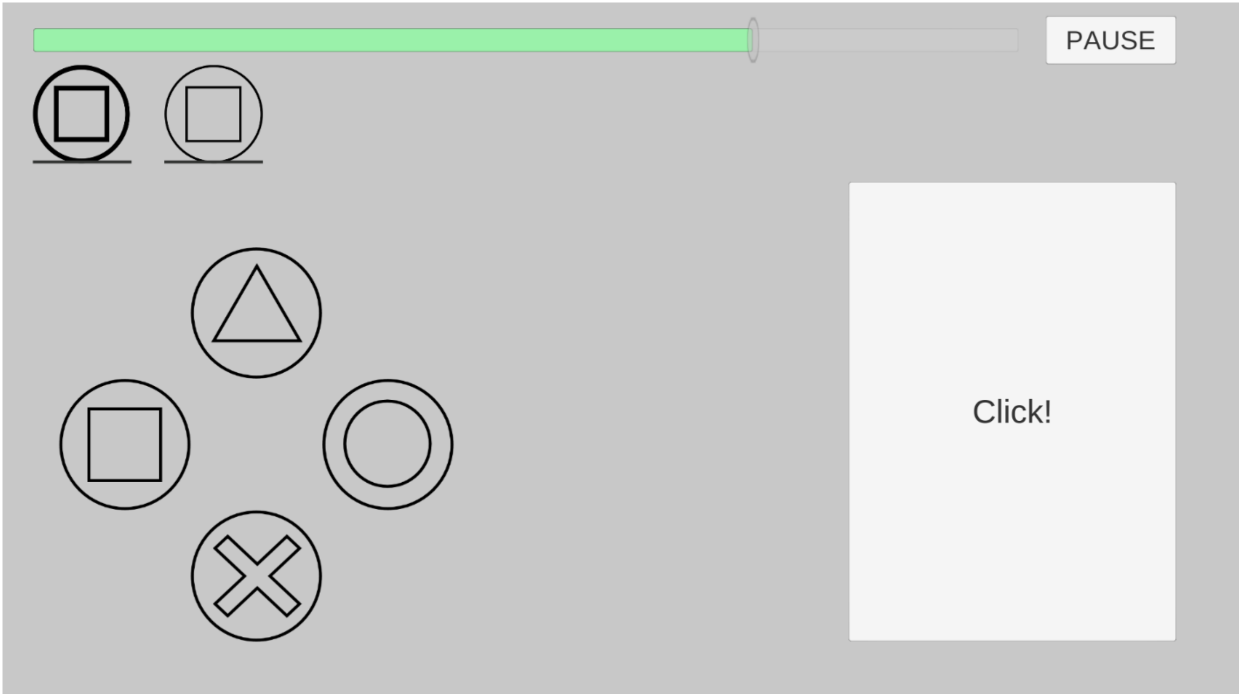


Figure 31 – Incorrect Input of First Shape

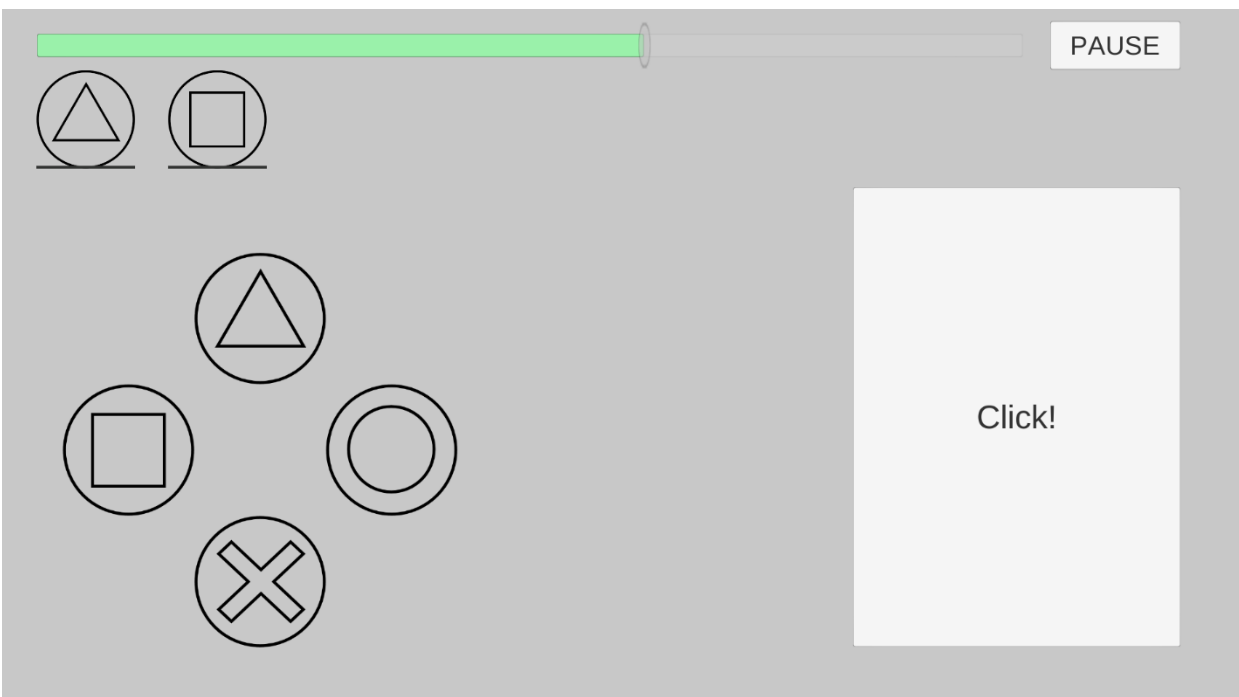


Figure 32 – Successful Input of Two Shapes

Figure 33 – Appearance of Stimulus Box

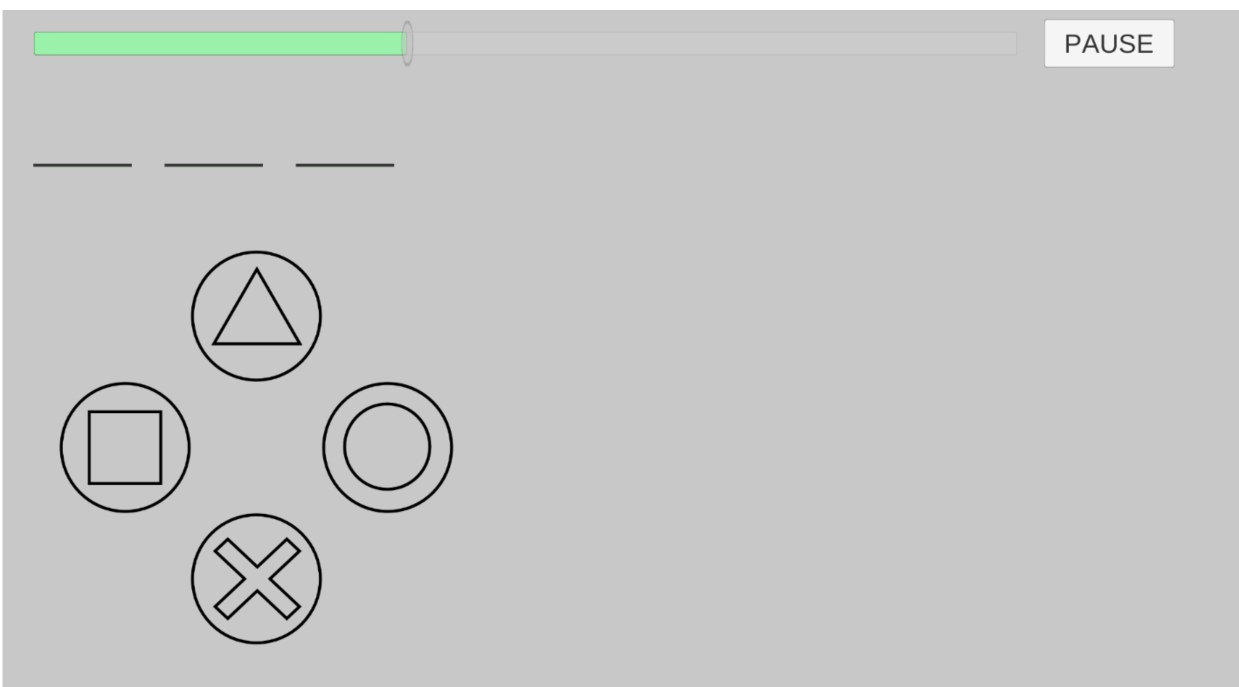


Figure 34 – Box Disappears After it Has Been Clicked