UPPSALA
UNIVERSITET

IT 20 007
Examensarbete 30 hp
Juni 2020

Anomaly detection in displays

Optimizing color coding for visual search

Sita Aukje Vriend

Institutionen för informationsteknologi
Department of Information Technology

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Anomaly detection in displays: Optimizing color coding for visual search

Sita Aukje Vriend

When monitoring complex physical processes, an operator often needs to make quick decisions based on real-time information. The information - often dynamic - can comprise many variables that need to be observed simultaneously. Graphical displays have perceptual properties that allow observers to quickly and accurately detect information and to observe the relationship between the many variables. This thesis implemented and examined chromatic perceptual properties of an integral display. Two color-coded integral displays were designed: one integral display with all states color coded (ID4), and one integral display color coded the dominant state (ID1). Two experiments compared the color-coded integral displays with a configural coordinate display: the first experiment examined static displays in which participants were tasked to identify a unique, known state; the second experiment examined dynamic displays with the unique state unknown to the participants. Both experiments revealed that the color-coded integral displays were more efficient than the configural coordinate display; the integral display with the dominant state color-coded was most efficient. Thus a color-coded integral display allows many variables to be observed simultaneously with high efficiency.

Handledare: Mikael Laaksoharju
Ämnesgranskare: Mats Lind
Examinator: Anders Arweström Jansson
IT 20 007
Tryckt av: Reprocentralen ITC

1 Acknowledgements

I would like to thank everyone involved who helped me to complete this thesis. First of all, I would like to express my gratitude to everyone who took the time to read this thesis and gave me feedback. I am also grateful for all the participants who took time out of their day to come to the lab for the experiments.

My special thanks goes out to Darren Lehane who provided me with his testbed and the data he collected during his experiments. To Mattias Kronberg for reading my thesis and being my opponent. To Anders Arweström Jansson for reading and examining my work. And finally, my special thanks goes out to Mats Lind and Mikael Laaksoharju for guiding me throughout this thesis and, especially, for helping me understand and structure the data.

Contents

Glossary

anomaly An unanticipated and unfamiliar event as defined by Vincente [39]. vi, 1–4, 7, 9, 13, 14, 16, 17, 23, 24, 34, 41, 43

configural dimension The dimensions perceptual identities remain unique and can be processed individually. The interaction between the dimensions creates new emergent properties [4] [40]. vi, 8, 13, 16

control task A task in which the observer needs to discriminate only one of two dimensions at a given time. The other dimensions value is held constant [4]. The use of parentheses is an example of a control task. When one of the parentheses is known, the other parenthesis is also known since they depend on each other. In this case, the side the parenthesis is facing is the dimension. vi, 7, 8, 13

divided task A task in which the observer needs to consider both (stimulus) dimensions to come to a conclusion [4]. vi, 7, 8, 13

effective In this thesis, effective was defined as the proportion of accurately identified display in an anomaly state. vi, 1, 2, 4, 6–9, 33, 42

efficient In this thesis, efficiency was defined as the latency in seconds it took participants to identify the display in an anomaly state from a set of displays. vi, 1–4, 7–9, 13, 16, 17, 28, 32–34, 37, 41–43, 48

emergent features The interaction between configural stimuli produce perceptual properties. The separate configural stimuli are processed by perception as a whole. Parts of a group of stimuli are perceived as a combination, and thus become wholes. Emergent features are inherently salient [4] [40]. vi, 8, 9, 13, 14

explicit attentional capture a salient stimulus that is not in focus draws attention. This leads to awareness of its presence [34]. vi, 2, 16

focused task A task in which the observer needs to keep track of variables of individual variable [4]. vi

human error An action made by a human that was not intended, not desired by an external observer or by a set of rules, or did not lead to a desired outcome. The human rarely is the sole cause of the error, as the error can often be traced back to faulty system design [31]. vi, 1, 2, 4, 9, 23

integral dimension The individual stimuli dimensions are not processed individually because of an interaction between the stimulus dimensions. Instead, its parts are perceived holistically [4]. vi, 8, 14, 16, 17

integrated task A task in which the observer needs to keep track of relationships across many variables [4]. vi, 7, 8, 16, 17

parallel processing When the understanding and processing of two different layers of variables is supported by visual displays [4]. vi, 8, 13, 14, 16, 28, 33, 34, 36, 37, 40–43, 48, 49

photoreceptors Specialized nerve cells also known as sensory receptors that are located in the back on the eye. These cells detect light-wave information that falls on to them. There exist 2 types of photoreceptors: cones and rods. vi

redundancy task Similarly to a selective attention task, the observer needs to consider only one of two dimensions. However, only one variable is presented in a particular condition so that the operator only needs to attend to either of the two variables [4]. vi, 7, 8

saccades The frequent and rapid movements eyes make to reposition vision and what the eyes look at. Saccades can be deliberate but often they are automatic and happen unnoticed [29]. vi

saccadic suppression A temporary visual anaesthesia during saccades in which visual sensitivity is reduced to particular stimulus configurations [29]. vi

saliency The extent to which a visual feature is outstanding compared to its environment and other visual features surrounding it. Visual salience can also be called visual prominence [4]. vi, 4, 5, 9, 16, 34

selective attention task A task in which the observer needs to consider only one of two (stimuli) dimensions. The second dimension will not remain constant [4]. vi, 7, 8

separable dimension There is no interaction between the stimulus dimensions. The perceptual identity of both dimensions remain unique and thus do not influence the perception of one another [4]. vi, 8

stimulus dimension The smallest part, feature, element or property of a visual stimuli. For example its shape, location, size or color [40]. vi, 5, 6, 8, 9, 13, 14, 16, 17

vigilance The ability to maintain concentrated and alert to a stimuli over a long period of time [42]. vi, 2, 27, 43

visual search The act of visually searching for a target item: objects such as people for example or a feature such as the color of an item for example, among other distracting items [25] [43]. vi, 5, 6, 28, 29, 32–34, 36, 37, 40–43, 48, 49

Acronyms

CCD Configural Coordinate Display. iv, vi, 1–4, 7–9, 11–14, 16–19, 25, 28–34, 36–46, 48, 49, 59

EID Ecological Interface Design. vi, 12–14

ID Integral Display. vi, 1–4, 7–9, 11, 13–15, 17, 18, 28, 32, 33, 41, 42, 44–46

ID1 Integral Display with dominant quadrant colored. iv, vi, 13, 15–17, 19, 25, 28–34, 36–43, 45, 46, 48, 49, 58

ID4 Integral Display with all four quadrants colored. iv, vi, 13, 15–19, 25, 28–34, 36–43, 45, 46, 48, 60

TMIM Three Mile Island Meltdown. vi, 1

2 Introduction

Not all decision-making in complex computer systems can be automated. In certain situations, a human is required to decide on a course of action once a system reaches a specific state. In other situations, it is necessary to display neutral information, as different states require a different course of action. Thus we need human operators to monitor these systems and intervene when necessary. The human monitoring such systems is required to make an educated decision as fast as possible, based on the information given by graphical displays. Therefore it is crucial that the human monitoring such systems can detect an anomaly effectively and efficiently.

Complex technological systems such as nuclear power plants, factories, etc. require operators to decide a course of action when the system reaches an unexpected anomaly state. Additionally, incidents such as the Three Mile Island Meltdown (TMIM), where human error was predominantly to blame, cause authorities to examine the fit between human and machine [33]. Human factors researchers studied the design mistakes of the TMIM control room to find layout design principals and suggestions for future control rooms. Display design was one of these aspects. How operators perceive and understand the complex technological systems they work with is one of the questions posed by human-factors engineers. Graphical displays can be considered a part of the answer to these questions, since they can effectively and efficiently communicate system states visually [6].

Investigation groups that analyzed the TMIM discovered that part of the problem was that the information the operators obtained at the time of the accident was ambiguous [33]. Complex technological systems should have displays that convey critical information, such as machine status, as effectively and efficiently as possible to the operators. This is why research that improves such displays is of importance, since they can prevent accidents such as at TMIM.

This thesis is built on previous thesis work that evaluated such displays [18]. Lehane's thesis compared the effectiveness and efficiency of different data visualization techniques for monitoring processes and anomaly identification on a dashboard-based user interface. Prototypes of three display types were developed: a standard bar graph, an Integral Display (ID) (figure 2) and a CCD (figure 1). The measured accuracy and response time hints that the CCD and the ID allowed for the most effective and efficient performance.

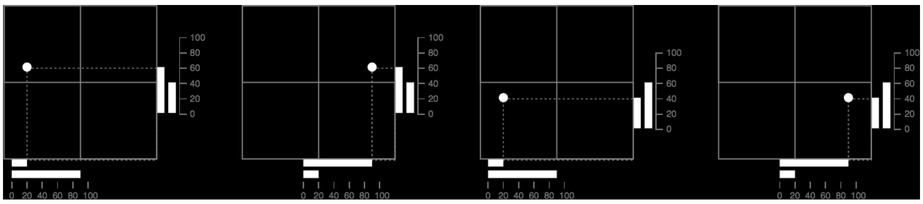


Figure 1: The CCD graphical display representing 4 dominant system states. The states read from left to right: 1, 2, 3, 4 [18].

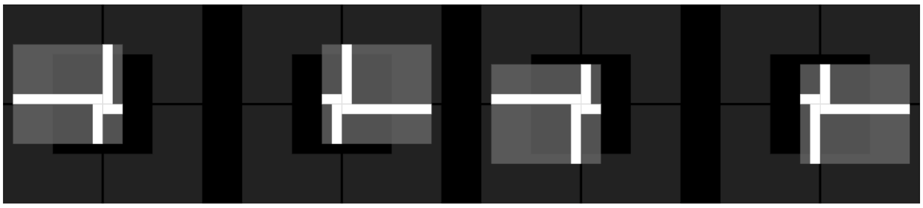


Figure 2: The ID graph designed by Laaksoharju and Lind, studied by Löfvenberg [20] and Lehane [18]. The dominant system states read from left to right: 1, 2, 3, 4

Since the previous study investigated configural properties of the ID, no color was used in the display. However, a recommendation was to add color and study the effects the addition color has on the perceptual properties of the display [18]. Therefore this thesis aimed to investigate how to add color to the ID and to evaluate the effects of the addition of color to the graph. The overall aim of this thesis is to further improve the efficiency of the ID.

Both the CCD and ID are proven to be effective and efficient at conveying the dominant state of a system [1], [3], [18], [20]. Only one state of interest is presented at any given time, which aids explicit attentional capture [34]. Criticism on research and development of displays like the CCD and ID could thus include the question: why should complex systems not just have automated response systems? However, while an automated response system would be beneficial in some complex systems, since it eliminates the possibility of human error, it is not a suitable solution for systems in which the response to a system state is unknown, or when multiple responses fit a single system state. The automation of any system aims to replace manual control, planning and problem solving of a human operator with computer systems [2]. Bibby and colleagues stated in 1975 that automated systems still require human intervention for supervision and maintenance [7]; it still holds true today. As Bainbridge stated, a human operator is required to monitor the automated system to check whether it operates as intended, and a human operator is required to take over tasks from the automated system in case it malfunctions [2]. On the other hand would an automated response system be suitable if a specific system state is related to one response that fits the state. Otherwise an operator is needed to make an educated decision. The system states of the complex systems should, in these cases, be communicated with efficiency and effectiveness. Displays such as the CCD and ID allow operators of complex systems to understand and diagnose the systems, so they can decide what response is most suitable according to the state.

An operator could be tasked to detect an anomaly in a system to check whether the automated system is malfunctioning for a long period of time. However, one of the challenges that comes with monitoring a system state is vigilance. It is known that it is difficult for humans to perform vigilance tasks for longer than 30 minutes [21] [42]. Displays that consider a suitable human-machine fit can help an operator quickly detect malfunctioning of an automatic system, and could help a human operator overcome their vigilance limitation during tasks such as these [2].

2.1 Aim

This thesis focused on improving the efficiency of the ID graphical display designed by Laaksoharju and Lind [20] and examined by Löfvenberg [20] and Lehane [18]. As Lehane suggested, this thesis investigated how to add color to improve the ID graphical displays. Possible solutions regarding the addition of color were researched, implemented and then evaluated in two experiments.

First, literature was consulted to find the best solution for adding color to the ID. Then the improved ID graphical display was implemented into Lehanes testbed to be evaluated during two experiments which evaluated the effect color had on the ID. The aim posed in this thesis was whether color improves the efficiency of the ID over the CCD. In this thesis, efficiency was defined as the latency it takes to identify an anomaly from a set. Thus, according to the aim should new ID designs allow for faster anomaly detection than the CCD.

> Thesis aim: *Improve the efficiency of the ID so that the ID becomes more efficient than the CCD.*

2.2 Limitations

Lehane's work focused on three detection tasks: known state detection, unknown unique state detection and state identification. The analysis of results from this thesis work is limited to *known unique state detection* and *unknown unique state detection* only. In addition, this thesis focused on investigating both static displays in which the unique state is known and dynamic displays in which the unique state is unknown. The investigation into dynamic displays in which the unique state is unknown would especially be relevant to human factor engineers and control room design, since complex systems are likely to require operators to analyze such data. Dynamicity has not received much attention in display research.

The developed ID displays of this thesis did not accommodate users who are unable to distinguish between the colors used in the developed display. It was considered likely that employers who will use the displays screen for color deficiency during job applications. Individuals with color deficiency are thus excluded from the evaluation study. Thus a screening and exclusion for such individuals was part of this thesis work.

3 Theory

The aim of both CCD and ID is to visualize the relationship between two variables. Both displays allow the operator to detect unique system states, detect specific systems states or identify system states, in both dynamic and static display types.

While automation of complex systems is often considered ideal, since it eliminates human error completely, it is not always feasible. Some systems are required to show neutral system states, since different decisions need to be made during different system states. Even when it is possible for a said system to detect the dominant state, automation might not be the optimal solution, since it is not always clear what the course of action should be once the system reaches a specific dominant state. Automation of complex systems is often suitable during normal situations. In abnormal situations, when a creative problem solver is required, controls are expected to be transferred to a human operator [15]. Thus, automatic systems need a human operator in case of system failure. In addition to the need for an operator to recognize such a failure, there is also a need for the operator to take over system controls [2]. In complex systems such as nuclear reactors, it is essential that the recognition of the system failure happens effectively and efficiently, so that system controls can be taken over without delay.

A system failure in complex systems can be classified as an anomaly: an unfamiliar and unanticipated event. An anomaly is considered the biggest threat to system safety, as such events are susceptible to human error [39] [15]. Examples of such anomalies include the malfunctioning of an automated system. Therefore, a suitable human-machine fit should be considered as a solution to support operators to identify unfamiliar and unanticipated events or anomalies quickly, so they can deal with the event in an appropriate manner.

The theory behind several related topics were explored in order to understand, and thus further develop, the ID display. Since the displays used in this thesis are visual, visual search was explored first. The development of new graphical displays needs to take into account how humans perform visual searches, how visual saliency can be improved, and how visual searches can be improved. Then the topic of automation from a human factors perceptive was explored. While the main purpose of automation is to eliminate human error, human operators are still expected to recognize system failures. Thus understanding how humans behave in automatic systems is important, since the ID could play a role in such complex systems. Both the ID and the CCD aim to support creative decision making in unfamiliar and unexpected situations in complex systems, such as automated systems. The theory from the topics mentioned formed a basis to explore display design and development. Here, design guidelines and examples were explored. Finally, research behind ID and CCD was explored.

3.1 Visual search

Since the area of the fovea is small, the detection field for the target item of a visual search is rather small. Therefore, multiple eye movements need to be made during a search task. The process of a visual search task can be broken down into three loops that are intertwined: the move and scan loop, the eye movement control loop, and the pattern-testing loop. When the target item is known, a search starts with body and head movements to get an optimal overview of the environment during the move and scan loop. From there, planning and making eye movements take place. Potential target items are judged for elementary features of the target item, such as size and color. Visual working memory keeps track of the recently visited locations of the eyes during this loop. Once a potential target item is found, the pattern-testing loop is executed. Here, the pattern is tested to check if the potential target item is the actual target item. Each pattern test takes around one-twentieth of a second. [41]

How quickly the target item is found in an environment depends on the difference in saliency between a target items features and the environment with other items. If the feature difference is big enough, the pop-out effect will occur. The target item will simply pop out and can quickly be recognized, since this effect is processed early on in the brain. Features that can aid the pop-out effect include the stimulus dimensions of color, elementary shape, motion and spatial grouping [36] [41].

The processing of sensory information by the brain can be divided into two forms of processing: bottom-up processing and top-down processing. Bottom-up processing is stimulus-driven, which means that it start at the bottom, with a visual stimuli, and ends at the top, where the brain understands the visual stimuli [43]. The pop-out effect is an example of bottom-up processing. Top-down processing, on the other hand, starts at the top, the brain, which tries to understand what happens at the sensory level [11]. The retinal image provided by the eye forms features in the brain, that are in turn translated into pattern, which builds the objects the brain can understand and interpret [41]. This type of processing is user-driven, which means that the user is consciously looking for the target item or feature [43]. Top-down processing is driven by the necessity to accomplish a goal and is led by attention [41]. Top-down processing is driven by cognitive processes, such as memories and expectations [11]. This can lead to biases that favor the signals one looks for during feature and pattern analysis [41]. Most visual perception and search tasks that are performed on a daily basis involve both top-down an bottom-up processing.

Treisman distinguished between two different search tasks: serial search tasks and parallel search tasks. In a serial search task all items need to be individually examined in order to find the target item. The time taken for this type of search task is defined by the conjunction of two stimulus dimensions. Therefore, the search time for the target item will increase when the set size increases. [37]

In a parallel search task all items are examined in parallel along a single stimulus dimension [37], also the target item. Here, a guiding feature defines the target [44]. As a result, the target item pops out for that single stimulus dimension, like for the pop-out effect. The pop-out effect is in fact a form of parallel search. The search time for a parallel search task is independent of set size [23]; a bigger set size will not slow down the time it takes to find the target item. A parallel search task can be guided by bottom-up processing.

Another type of search tasks, which lie in between parallel and serial search tasks, are guided search tasks. Guided search tasks were defined by Wolfe as visual search that allow for filtering out a subset of possible target items, according to basic visual features. Within a set of items, the attention is restricted to a subset that match basic features of the target item. First bottom-up processing filters out the subset that contains the target item, then top-down processing examines the remaining items. [43]

3.2 Color & color perception

A color can be described by 3 features: hue, saturation and brightness. Hue is understood as the wavelength of the light of a stimulus. Saturation describes how deep and pure the color is, and brightness describes how much light is emitted or reflected by the stimulus [11].

However, a color is never seen in isolation. The perception of a color is affected by the colors surrounding it [41]. One factor that influences the perception of a color is contrast, which describes the difference in brightness or color of a stimulus. A high contrast allows visual features to be distinguishable from one another. In addition, contrasting colors have the ability to attract attention [35].

3.2.1 Decision making in automated systems

Operators monitoring automated systems for anomalies require creative problem solving skills, as it might not be clear what the appropriate course of action is. Knowledge of problem solving and decision making is important for the design of effective displays, as they map the work domain or problem space to the display, and from the display to the human operator. Furthermore, a deep understanding of the work domain or problem space is important in the development of effective support systems, such as visual displays. The work domain will add a set of constraints such as tasks, goals and limits that aid the development [5].

Besides a good understanding of the work domain, is it important to identify the abilities and limitations of the operators who monitor these complex systems, in order to design suitable decision support systems such as displays. Development of such systems should especially recognize the human information processing capabilities, in such a manner that does not overload the operator. Such limitations can, in turn, facilitate the development of performance aids that would lead to safer and more effective system performance [15].

Both the ID and the CCD were designed to support the operators decision making during anomaly conditions. The operator needs to recognize and diagnose efficiently the anomaly before deciding on an appropriate course of action, such as monitoring and controlling system resources, selecting between alternatives, revising diagnoses and goals, determining the validity of data, overriding automatic processes, and coordinating the activities of other people [15].

A model of problem solving developed by Rouse [30] assumes the human preference for pattern recognition for the solutions of problems. This model suggests that problem solving happens on three levels. The first level is recognition and classification, the second level is planning, and the third level is execution and monitoring. Displays such as ID and CCD should support the first and third level of this model to be effective and efficient.

3.3 Displays

Since automated systems tend to produce a high number of false alarms in unfamiliar and unexpected events [28], human operators are still required to observe these systems and act as creative problem solvers in case of an anomaly. Displays that are effective and efficient aid operators of complex systems with decision making and problem solving during anomaly events [5].

While auditory and tactile displays exist and can be beneficial in work domains that employ complex systems, this thesis focused on visual displays only. Visual displays can be divided into two groups: static displays which include for example road signs, and dynamic displays which change over time and include for example speedometers. According to Bridger, good visual display design allows for integration of information and its design depends on the task an operator has to perform [9].

3.3.1 Design

Bennett and Flach define a graphical display as a system that *"maps information from a domain into visual features"* [4]. A display in a human-machine system is a representation of the underlying domain. The operators tasks ought to be defined by the work domain and not by the visual characteristics of the display. Bennett and Flach further state that humans perception-action skills such as their pattern recognition capabilities should be taken advantage of, and powerful interface technologies should be used when appropriate [4] [5]. Thus both work domain and operators tasks, and perception skills should be considered when designing a display.

Bennett and Flach [4] sketch out important properties which graphical displays should have. These properties depend on the task, since any display needs to suit the work domain. Bennet and Flach described visual discrimination tasks which operators of complex systems might need to carry out [4]: redundancy task, divided task, selective attention task, integrated task and control task. In a control task or baseline task, only one of two stimulus dimensions need to be considered to make the discrimination. The other stimulus dimensions value will be held constant. The use of parentheses is an example of a control task. When the right-facing parenthesis is discriminated, the left-facing parenthesis is held constant. When one of the parentheses is known, the other parenthesis is also known since they depend on each other. In a selective attention task of filtering task the discrimination depends on one of two stimulus dimensions, similarly to a control task, but none of the stimulus dimensions are held constant. A redundancy task or correlated task requires the operator to make a discrimination by either one of two stimulus dimensions, or by attention to both stimulus dimensions. An integrated task is a task where the operator is required to track the relationship between stimulus dimensions. Finally, a divided task or condensation task is one where the operator needs to consider both stimulus dimensions to make the discrimination.

The stimulus dimensions an operator uses for any of the mentioned tasks are visual stimuli such as shape, location or color. These stimulus dimensions are divided into three categories: separable dimension, integral dimension and configural dimension. A separable dimension means that there is no interaction between stimulus dimensions. The stimulus dimensions remain independent and do not influence the perceptual identity of one another. No emergent features can arise from the interaction between the stimulus dimensions, since no interaction exists. Separable dimensions are processed independently of one another. A task that benefits from separable dimension is a selective attention task. An integral dimension is one that sacrifices the perceptual identities of individual stimulus dimensions to show the interaction between the two stimulus dimensions. A redundancy gain emerges from the independent processing of the two stimulus dimensions. Tasks that benefit from integral dimension include redundancy task. A configural dimension is a compromise between integral dimension and separable dimension, since it can be processed both independently and interdependently. While both stimulus dimensions remain perceptually unique, emergent features arise from the interaction between the two stimulus dimensions. A configural dimension is suited for divided task and control task [4].

Operators of dynamic systems often have to make decisions based on both high-level constraints (e.g., status of processes) and low-level data (e.g., output of sensors) [4], a divided task. The stimulus dimension best suited for such a task is configural dimension since it allows for parallel processing since parallel processing is a divided task.

Displays that are designed specifically for divided task are configural displays. These reduce divided and focused attention demands. Bennett and Flach stated that *"configural displays provide an opportunity to turn data into information by presenting those data in a context that reflects the process constraints."*. The CCD and ID displays are examples of configural displays that have been proven to allow for the processing of large quantities of data with a low attention cost [18] [20]. For both displays, it is important that both the system states, a high-level process, and the individual variables, the low-level process, can be interpreted by the operator effectively and efficiently.

To reduce attention cost further, effective and efficient displays can take advantage of principles that aid human visual perception to increase the saliency of certain stimulus dimensions. This can be achieved by improving bottom-up processing, for example, by implementing the pop-out effect [36]. Use of emergent features also seems to enhance saliency [4].

When developing displays, it is important to distinguish between design for data availability and design for information extraction. When data availability is only considered, operators are left to collect the relevant data, maintaining the data in memory and then mentally integrating the data in order to make a decision. This can lead to poor decision making and human error [45].

The ID and CCD have both been proven to be suitable for anomaly detection [18]. Both incorporate concepts from visual search and visual design as mentioned in this and earlier sections.

4 Testbed

The original software testbed was developed by Lehane [18] for the purpose of evaluating performance of different display types that visualize data of a shared data set. The testbed is a custom-made web application that allows for simultaneous multi-user testing. Participants can access the software via a web browser, as the testbed is hosted remotely. The participants' results are recorded to a separate PostreSQL database after the participants complete a screen of the study. This testbed is used for the current study, since it builds upon Lehane's work and since it suits the purpose of the current research. However, changes needed to be made to serve the needs of this study. An overview of the layout of the testbed used in this thesis can be found in appendix E.

4.1 Colors

The aim was to improve performance by increasing the saliency of the parts of the ID that convey the important information, and to highlight the difference between the parts. The information should be conveyed more efficiently and effectively. In other words, the addition of color was to improve the accuracy and response time of the display.

The original ID design has a dark grey background on which the dynamic data is presented in white, which is close to the maximum contrast that can be achieved (figure: 2). However, the original displays only use one physical form: contrast. The addition of color allows for redundancy gain, which states that additional physical forms or stimulus dimension such as the addition of color can allow better efficiency by alternative physical forms [32]. Another function of adding color to data representation aims to make the four different states distinguishable from each other, but not to make one state more salient than the other states. Furthermore, literature suggests that more saturated colors provide stronger signals on one or more chromatic channels [41].

As previous work suggested [18], four suitable colors were chosen to represent the four different states. A trade-off was made between providing a strong signal for each state and to maximize the contrast with the background. The colors were to be as light as possible, while still distinguishable from each other. Thus, the hues should vary while the perceivable brightness should remain the same per color, and should lay around 50% between black and white. The choice of hues did not have to take colorblindness into account. The ideal hues are therefore red, green, blue and yellow. While there exists an overlap in the cone-sensitivity functions, the human eye is sensitive to red-green differences. This is because color-opponent theory states that the red-green channel is the difference between the signal from the middle- and long wavelength-sensitive cones [10].

A formula for finding the brightness of an RGB-color is [26]:

$$brightness = 0.21 * R + 0.72 * G + 0.07 * B \tag{1}$$

Since the maximum values of R, G or B can not exceed 255, the brightness cannot exceed 255. The chosen hues have to contrast against white and black, they should have a brightness of 50% between black and white. Therefore $brightness = 128$.

$$128 = 0.21 * R + 0.72 * G + 0.07 * B \tag{2}$$

The blue hue was found by maximizing B and minimizing R: $B = 255; R = 0$. G was found to be 152, therefore, the blue hue was set to RGB(0,152,255).

The red hue was found by maximizing R and minimizing B: $B = 0; R = 255$. The equation resulted in $G = 1$, which made the red RGB(255,1,0).

The green and yellow colors were calculated differently. For the yellow color, an attempt by trial and error was made to find a color with equal red and green, by minimizing blue. The yellow color turned out to be RGB(138,138,0) and was discarded since it was considered too similar to the green hue.

The yellow hue was substituted with RGB (255,78,255), a magenta hue that was considered distinguishable enough from the other chosen hues. This hue was found by looking for a hue that lies between red and blue. Both blue and red were maximized: $B = 255; R = 255$.

The green hue was calculated by setting green to 177 and red to 0: $G = 177; R = 0$. This resulted in the green hue to be RGB(0,177,0).

A change in saturation based on the squares size was proposed. This added another limitation in which the chosen hues should be of the same saturation.

$$S = \frac{max - min}{max + min} \tag{3}$$

Preferably, S should be as close to 1 as possible: $S \approx 1$. This proved difficult, since all hues except the magenta hue had at least one digit set to 0. The saturation for the magenta hue was $S = 0.5$, while it was $S = 1$ for the green, red and blue hue. While the saturation of all hues did not match, it was decided to keep the current hues due to time constrains. A table with the properties of the four final colors that were chosen can be found in table 1.

Name	R-value	G-value	B-value	Saturation
Blue	0	152	255	1
Red	255	1	0	1
Green	0	177	0	1
Magenta	255	78	255	0,5

Table 1: Properties of the chosen colors.

4.1.1 Color pre-test

It was assumed that employees working with such systems as the CCD and ID should be able to distinguish between the colors that were decided upon. Therefore the experiments exclude individuals who have a color deficiency during participant recruitment.

First the Ishihara-test for color deficiency was considered to filter out possible participants with any color deficiency [17]. However, participants who had a color deficiency were not considered a problem, it is only important for participants to be able to distinguish between the chosen colors. The color pre-test was both created to be a stand-alone web application and to be added to the testbed.

Figure 3 and 4 show the color pre-test. Since the colors were shown for a very short time, it was important that participant was ready and prepared to spot the colors. Therefore participants had to press the spacebar to show the two colors every time. Once the colors were shown and the squares were black again, the participant had to press the left arrow key if the colors were the same, or the right arrow key if the colors were not the same.

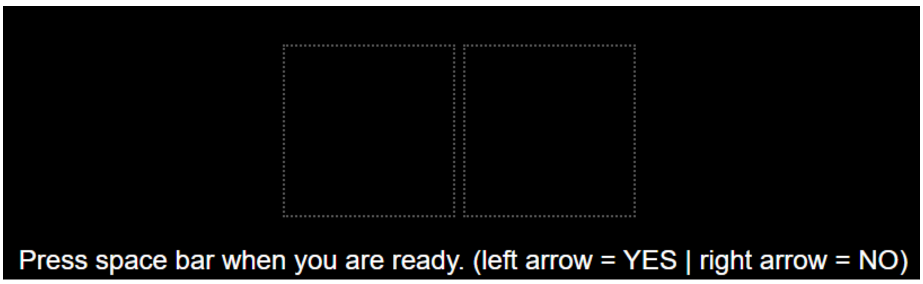


Figure 3: The color pre-test before the colors were shown. The spacebar is pressed to show the colors as in figure 4.

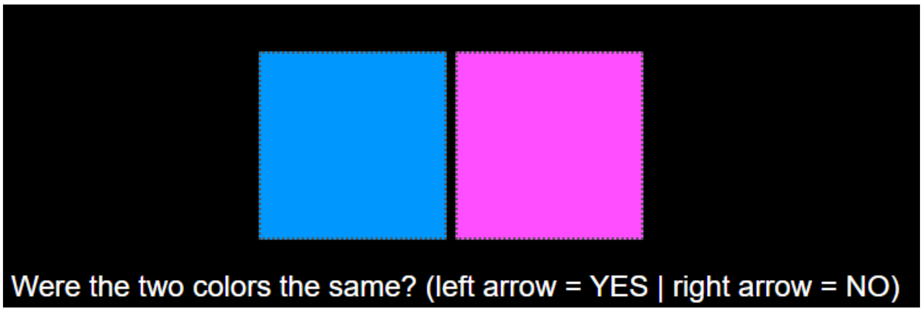


Figure 4: What was displayed for a very short time after the spacebar was pressed. When the squares become black again as shown in figure 3, the participant is asked to press the corresponding arrow button if the colors were the same or not.

A link to the stand-alone web application was sent out to potential participants for pre-screening. The pre-test at home was considered discouragement for those who failed it and practice for participants who succeeded. Upon completion and scoring 90% or higher correct, participants were directed to a form where they could sign up for the study. Potential participants who failed the pre-test were not directed to the form and discouraged from signing up. However, no data from the pre-test was stored in an online database and participants were able to perform the pre-test multiple times. It prepared potential participants for the pre-test in the lab. Participants who would not have been able to distinguish between the colors would likely not have passed the color pre-test in the lab.

The pre-test was added to the testbed after the screen where participants received their unique participant-id number for all display types. This was before the participants began the actual experiments of the different displays. Participants had to do the color pre-test again to control for potential confounds that might have occurred when performing the same test in a home environment. For example, if participants were able to perform the pre-test as often as they liked, they could ask someone else to take the pre-test for them, and is it likely that participants had different monitors with different color accuracies or brightness settings. Since the participants were to be aided and not hindered by the colors, they were supposed to be able to distinguish the colors on the same screen as the screen on which the evaluation took place.

4.2 CCD

The CCD-display (figure: 5) is based upon the Ecological Interface Design (EID) framework developed by Rasmussen and Vincente [39]. The framework is a use-centered approach to display design [14] that adds the work domain, or ecology, to the human-technology relationship. The EID framework is used to gain understanding of the work domain of complex work systems, since this framework is suitable for development of controls of different work domains to help operators cope with unanticipated events [39]. Its aim to aid design for operators to deal with the complexity of a system, by considering the work environment or ecology in which the operator and machine work together [8].

The EID framework has been applied in a wide variety of work domains for more than twenty-five years [1]. Examples can be found in interfaces for a milk pasteurizing plant [38], power plant [12], marine visualization [28], nuclear power plant [1] and an intelligent human-machine system [22]. Thus far, research of the applications have shown that an ecological interface such as the CCD is an efficient tool to support operators in recognition of complex and unusual situations and in decision-making [3]. The ecological framework approach to visualization information improved situational awareness and supported the decision-making process, especially in unfamiliar and unanticipated complicated situations, or anomaly events [1].

The CCD display was designed to visualize the relationship between variables and system states. The spatial location of a dot on the coordinate grid represents the system state, which was designed to be an emergent features [16]. The main stimulus dimension used is the dot location. In CCD, dot location is a configural dimension since both individual values and system state can be interpreted. The values from the x-axis and y-axis remain perceptually unique, while the interaction creates an emergent features that can be interpreted on its own. The CCD is best suited for divided task and control task.

When the CCD was compared against an alphanumeric display, a polar graphic display and a bar graph display, it was found that it performed best overall [16]. Configural displays such as CCD have been shown to engage parallel processing. They allow operators to process large quantities of data with little attention as both divided and focused attention demands are reduced [4]. While the CCD is easy to learn initially, since the dot is in the location of the system state and thus no mapping needs to be made, it does not allow for parallel processing in a situation where a large quantity of CCD displays are presented. When multiple graphs are shown on screen, a serial search needs to be conducted to find the anomaly state. This means that CCD is less suited when presented in large set sizes.

Adding color to the CCD was considered to allow for a fair comparison between the ID displays and the CCD. First, coloring the dot was considered, since that would be similar to ID4 and ID1. This change was dismissed as it was anticipated to decrease performance.

Adding color to the bars that are located below and to the side of the CCD, which was suggested by Lehane, was also considered. However, this change was thought to be too different from the improvements made in both versions if the ID display. Coloring the bars would not emphasize the state the display is in. Thus, the CCD was not changed and remained as implemented by Lehane, and as shown in figure 5: monochrome [18].

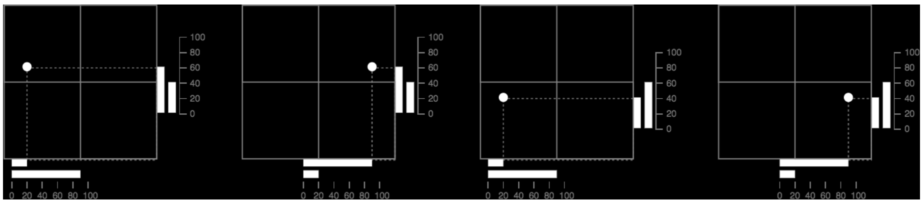


Figure 5: The CCD graphical display representing different dominant system state. The systems states as shown from left to right: 1, 2, 3, 4.

4.3 ID

The ID (figure: 2) is a poietic redesign of the CCD. The display was designed by Laaksoharju and Lind, presented by Löfvenberg [20], and then further developed and tested by Lehane [18].

Poietic design deals with information visualizations and allows users to gain a more intuitive and deep understanding of how the system works. According to poietic design, the artifact should be designed to show what it can do and what a human operator can do with it. The user should be able to form a mental model of the system from the design. This is achieved by making the characteristics of the artifacts visible. The human operator should interpret and judge the data from the artifacts. According to Lövfenberg [20]: *"The artifact should not judge these effects; it is rather up to the user to interpret the data."* Like EID, poietic design also considers the human-machine relation, as the framework takes a user-centered approach. The framework aims to translate the imperceptible into perceptible data.

The ID relies on the center of gravity of the graph to be perceived as a whole. This gravity concept is an integral dimension. The perceptual identities of the individual stimulus dimensions are sacrificed to show the interaction between the two stimulus dimensions: the emergent property or emergent features.

Löfvenbergs thesis proved that displays designed according to poietic design performs better than CCD for tasks the CCD was not designed for [20]. Lehanes thesis proved furthermore that the poieticly designed ID performed better than the CCD designed according to EID in anomaly detection regarding system states.

A learning curve might exist for ID, since a mapping needs to be made between features in the display and system states. The operator needs to judge the center of gravity and map that to the relevant system state. Once practiced with or without modifications that exploit the theory mentioned in previous sections, the ID could allow for parallel processing. This would mean that the search time to detect an anomaly would be independent of the set size.

Lehane [18] suggested that color could improve the ID display. He provided two potential designs: figure 6 and figure 7. Both designs assign a specified color to each state, and thus focus on enhancing the recognition of the dominant state.

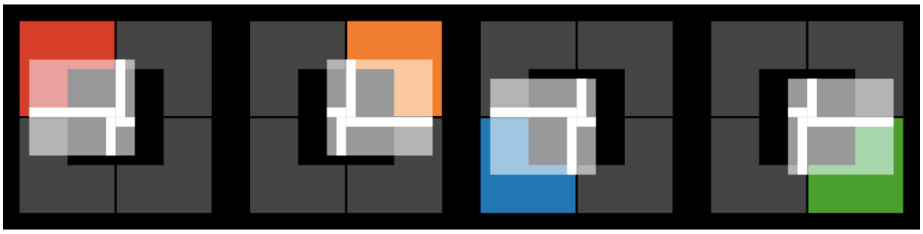


Figure 6: Possible design with color for ID according to Lehane [18]. The square in the background of the corresponding state is colored when that state is dominant. Each state has its own unique color.

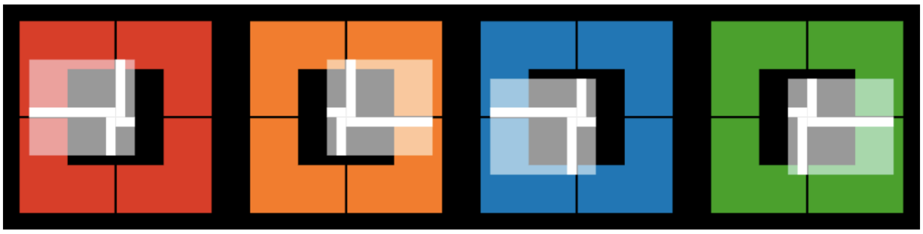


Figure 7: Design with color for ID. The entire background is colored with a unique color that corresponds to the dominant state.

Four possible parts of the ID display could be colored: (1) the whole background, as in figure 7, (2) the background quadrant related to the two variables, as in figure 6, (3) the bars, (4) and the square between the bars. The idea of coloring the whole background with the color related to the state was discarded, since it was considered overwhelming. This was also considered to draw the user's attention to the background rather than to the relevant information presented by the graph: the relationship between the variables and the dominant state. Coloring the background quadrant related to the state was considered better, since it would draw attention to the relevant information regarding the state. However, it would draw attention to the background rather than to the graph. Coloring the bars was dismissed due to the fact that it would not highlight the relationship between the variables. It was decided that the square between the bars was the best part of the graph to be colored, if that part was not transparent, as transparency allows for a smaller range of hues between a colors full saturation and full saturation. The square represents the relationship between the two variables; adding color to this part of the graph would draw attention to this relationship, which is the important information the graph should reveal. Additionally, it was decided to remove the transparency of this part of the graph to delimit the brightness and saturation constrains of the chosen hues.

Two new designs were created for the ID: the ID4 which shows all colors (figure: 9), and the ID1 which only shows the color of the dominant state (figure: 8). The graphical display that shows all colors (figure: 9) can be considered unbiased and would be beneficial when all system states are regarded as equally important. In this case, different states require a different course of action that the operator needs to decide upon. The graphical display that only shows color on the dominant state (figure: 8) aids explicit attentional capture. This graphical display shows all other states were shows with a grey hue and the dominant state using a specific color related to the state. It aims to make the dominate state more salient and conveys in roughly the same manner as the CCD.

4.3.1 ID1

The ID1 graphical display (figure 8) was designed according to the pop-out effect [36]. The display only shows the color of the dominant state of the system at any given moment. The other states have a grey color. In the dynamic condition, the state changes color gradually from the most saturated version of the states color to grey when the dominant state is shifting. The state which becomes dominant changes color from grey to the color of that given state. Since the ID1 aids explicit attentional capture, it was assumed that this graphical display was the most efficient graphical display overall.

The ID1 is designed to allow operators to spot an anomaly in large quantities of data, thus supporting parallel processing. Such a task can be considered an integrated task. Therefore the dominant, and thus relevant, state for this task, is made to be the most saliency. When many displays are presented, the anomaly state simply pops-out since its color is unique among the displays. As designed, the stimulus dimensions in ID1 could be considered not completely configural or integral. Coloring the dominant state can be considered to lean more towards an integral dimension. However, the size of the box of the dominant could be processed separately, allowing this stimulus dimension to be viewed as a configural dimension. The configural dimension is however less prominent than CCD, and it is unlikely that the individual X and Y values can be processed at the same time as system states.

Figure 8: The ID1 graphical display representing different system states. When dynamic, the last dominant state changes gradually to grey, and the new dominant state changes gradually to the color its state represents. The system states as shown from left to right: 1, 2, 3, 4.

4.3.2 ID4

The ID4 (figure: 9) shows the colors representing each system state consistently. While this graphical display was estimated to be less efficient than the ID1, it was estimated to be as efficient as the CCD. Because the ID4 shows unbiased system states, it could be beneficial for tasks where all system states are equally important. For example, when an operator needs to keep track of both the dominant and smallest states. While the current thesis would not evaluate the graphical displays on such a task, the ID4 was still included to examine whether the ID graphical displays are as efficient as the CCD.

Similar stimulus dimensions are used in ID1 as in ID4, and thus could not be considered completely configural or integral. The stimulus dimensions used are overall leaning more towards integrated task, since the area between the individual x and y values is more salient than the individual x and y values themselves. The individual x and y values can still be processed by judging the height of the white bars that lay on the axis, and by judging the height and width of the boxes. It is thus hypothesized that processing the individual x and y values at the same time as the system state is difficult. However, system state is the main task studied during the current thesis. An integral dimension such as the color and size of the boxes is suited for such a task.

The task studied in the current thesis was anomaly detection. The ID1 was hypothesized to be better suited. The dominant state is less salient among the other states within the display. Thus the dominant state does not pop-out. Guided search might be supported in ID4 when the anomaly state is known. The operator only has to focus on a subset of displays where the color of the anomaly state is large enough to be processed. Then a serial search is required to judge the area of the color for each state in the subset.

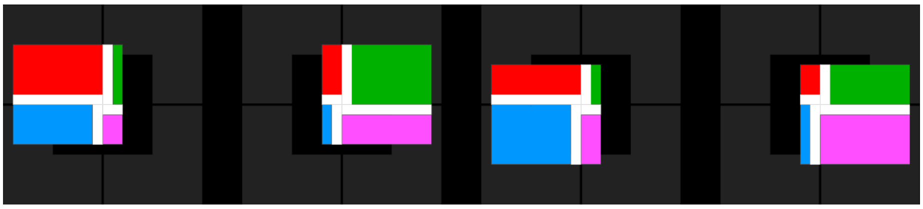


Figure 9: The ID4 graphical display representing different system states. The systems states as shown from left to right: 1, 2, 3, 4.

5 Data collection

This thesis performed two separate experiments: an experimental study (section 6) which focused on supporting the existing theory, and an exploratory study (section 7) which focused on the displays applicability in a more general setting. These two experiments used the same testbed (see section 4) and data collection methods, which are described in this section.

Both experiments aimed to evaluate the effects of color on the ID display. Both the ID4 and ID1 were analyzed for unique state detection tasks and specific state detection tasks. During the result analysis, both ID4 and ID1 were compared against CCD. The ID1 was hypothesized to be more efficient than the

CCD, and the ID4 was hypothesized to have a similar efficiency as the CCD. The aim related to these claims is: *Improve the efficiency of the ID using color so that the ID becomes more efficient than the CCD.*

5.1 Pilot studies

Two separate pilot studies were run to check whether the evaluation set-up and testbed were working properly, and whether all information given was clear enough to perform the evaluation without questions.

5.1.1 First pilot study

The first pilot study was done with a researcher from the department as a participant, who had knowledge of the project. The pilot was performed using the ID4. The goal of this first pilot study was to fix major hiccups that could hinder the participants during the actual evaluation. This included unclear instructions and bugs in the testbed.

No verbal instructions were given before the evaluation started. The pre-test questionnaire (appendix C), post-test questionnaire (appendix D) and the informed consent form (appendix A) were placed near to the computer, together with a pen. The participant was confused by the lack of verbal instructions before the evaluation started. It was also considered better to give the informed consent form before being placed in front of the PC. Also, the participants were given instructions to read the information on the screen carefully, as it contained all the information they needed.

A bug was found in the color pre-test where the button was located too low on the screen. The participant could not find the button without scrolling down the page. It was considered confusing, and the button was moved higher up so scrolling down was no longer necessary.

Instructions regarding what to do during the specified state task were also considered confusing. Thus the instructions were rewritten to clarify how the task was supposed to be performed.

5.1.2 Second pilot study

The second pilot study was done to approximate how long a participant would take to finish what was hypothesized to be longest condition: the CCD display. This display was assumed to take the longest time for participants to complete. This second pilot study used the same materials as used in the first pilot study. Only the hiccups found were resolved. The pilot study participant was an HCI master student. The participant was timed roughly from starting pre-study survey (appendix C) until finishing the post-study survey (appendix D).

The pre-test survey was confusing to the participant as it asked for two id numbers: one that was called participant number and one that was called *participant number - 72 charts*. The instructions on the unique id screen were thus changed to instruct the participants to leave *participant number - 72* charts blank until later instructed.

The pre-test to test for color deficiency had clear instructions, according to the participant. The participant had not performed the pre-test at home as all other participants were asked to do, but passed nonetheless, with few misclicks.

During the task that asked participants to identify sequential charts, it was noticed that the numerical pad on the keyboard was not turned on. When the numerical pad was turned on, it was not working. During the set-up at the evaluation, it was tested on several PCs whether the numerical pad worked. It was then determined that the numerical pad was not working. All participants used the number row above the letters on the keyboard.

The specified state task instructions page did not mention that the displays would be covered by a grey box from the beginning of the task. It was also not stated that the display could be uncovered to begin the task by pressing the Ctrl key. While the pilot participant managed to understand what to do, it was still considered confusing. The instruction page for the specified state task was updated to instruct participants that the charts would be covered by grey boxes at the start of the task, and that they should press Ctrl when they were ready to begin the task and to uncover the charts. It took the participant approximately 50 minutes to complete the evaluation.

5.2 Independent variables

The different independent variables that were tested included **display type, task, number of displays and dynamicity**. Participants tested combinations of these independent variables. An example of static 8 dis-plays with a known unique state for the ID1 can be found in appendix F. Appendix G shows an example screenshot of static 32 displays with a unknown unique state for the CCD. Finally, appendix H an example screenshot of static 32 displays with a known unique state for the ID4.

Display type: The type of graphical display the participants tested. Three different graphical displays were tested: *ID4, ID1 and CCD*. This condition was tested between subjects; participants were assigned to one of the three display types, and tested only on that one display.

Task: Which task participants were asked to perform. There were three tasks participants performed: *sequential display, known unique state, and unknown unique state*. *Sequential display* asked participants to identify which state a display was in by pressing the number keys 1 to 4 on the keyboard for each display on screen. This learning task that taught participants to understand and recognize the correct system states in the display. As this task was considered training, participants received feedback on whether they correctly identified the system state of each individual display. During the task *known unique state* participants were asked to identify which display was in a state that was unique compared to the other displays. Participants were told which state the unique display was supposed to be in. This task had a training phase in which participants received feedback on whether they identified the correct display, and a performance phase with no feedback. For the participants this task was called: Detecting a specified state. The task *unknown unique state* also asked participants to identify which display was in a state that was unique compared to the other displays, but did not name which state was going to be unique. This task had only a performance phase. This variable was tested within subjects, all participants performed all tasks.

No. Displays: The number of displays were shown on a particular screen. Three different number of screens were evaluated: *8-display, 32-display and 72-display screens.* This was tested within subjects, all participants had to test *8-display, 32-display and 72-display screens.*

Dynamicity: Whether a display was *static or dynamic.* The *dynamic* displays had variables and system states changing over time. The *static* displays showed an unchanging system state and variables. All displays on any particular screen were either static or dynamic. This condition was tested within subjects, all participants evaluated both static and dynamic displays.

5.2.1 Order of independent variables

The order in which participants performed tasks with the different combination of independent variables has remained as much as possible the same as previous work [18]. Some changes to the order were made for the benefit of both experiments. See table 10 for the order of the combination of independent variables used in this thesis.

After participants completed the color pre-test, they started with first trainings task (*sequential display training*), which was taught participants to correctly read each system state. The following instructions were given before the first training: *You will now see sets of charts whose states you must identify in sequential order, starting from top-left. When you believe you have identified which state a chart is in (1, 2, 3, or 4), you should press the corresponding number key on your keyboard. After pressing the number key, you will receive feedback on your answer.* Either 8 or 32 displays were shown on screen. Participants had to press the number key associated with the selected display. The selected display had a white box around the display. As soon as the participant pressed a number key, feedback would be given below the display and the next display would be selected. The feedback given told the participant if they were correct or not and in which state the display was in.

The second training task (*training - known unique state*) explained participants how to read the displays and their task before the training started. Participants received the following instructions on the task to perform: *You will identify the chart that is in a state that is specified above each set of charts.*

When entering a new page, all charts will be covered by a grey square. Read the task and state you are asked to find, these are located above the grey boxes. When ready press Ctrl to uncover the charts and begin your task. When you identify the correct chart, press the space bar. All of the charts will then be covered by a square. Use the mouse to click on the square that is covering the chart in the specified state you just identified. During training, participants received feedback on whether the correct display was chosen.

After the training with the static known display task, participants were directed to the performance phase (*performance - known unique state* and *performance - unknown unique state*). The instructions given were: *You will now complete a similar set of unique chart tasks again, but you will not receive feedback on your answers.* Participants performed this phase for screens that showed 8, 32 an 72 displays. For each number of displays shown on screen they repeated the task 4 times.

Participants ended the study with tasks with 72 display on screen (*training - known unique state, performance - known unique state and performance - unknown unique state*). Previous work reported that participants likely experienced fatigue towards the end of the evaluation because the original study took the participants longer than this study [18] [20]. The original study was considered to be too long and likely exhausted participants to the extent that it would have negatively impacted their accuracy and efficiency. Therefore, all 72-display conditions were placed at the end of the evaluation. In case of exhaustion, the 72-display would not compromise the efficiency and accuracy of the 8-display and 32-display conditions. This would also allow the 72-display conditions to be discarded in case the participants would drop accuracy significantly during these displays. The order of the combination of the dependent variables and how many times the participants tested these combinations (No. screens) can be seen in figure 10.

Task	No. Displays	Dynamicity	No. Screens
Sequencial Display Training	8	Static	1
		Dynamic	1
	32	Static	1
		Dynamic	1
Training - Known Unique State	8	Static	4
		Dynamic	4
	32	Static	4
		Dynamic	4
Performance - Known Unique State	8	Static	4
		Dynamic	4
	32	Static	4
		Dynamic	4
Performance - Unknown Unique State	8	Static	4
		Dynamic	4
	32	Static	4
		Dynamic	4
Training - Known Unique State	72	Static	4
		Dynamic	4
Performance - Known Unique State	72	Static	4
		Dynamic	4
Performance - Unknown Unique State	72	Static	4
		Dynamic	4

Figure 10: The order in which the participants tested the conditions in the experiments.

5.3 Dependent variables

Only data from the performance phase was used for result analysis (see figure 10). The training's phase was considered a learning phase for the participants in which they learned how identify each state and get acquainted with the controls.

5.3.1 Accuracy

The participants received the following instructions during the training phase:

> *"It is important that you try to be as accurate as possible. This is your main task. However, while being correct, you should be as fast as possible."*

Accuracy, in the context of this thesis, meant that the participant selected the display in a unique state different from the other displays on screen. The testbed determined on screen load which display would be in the anomaly state or enter the anomaly state. The states the other displays on screen would be in or would enter was also determined on screen load. However is was still possible for the other displays the to enter a unique state before the predetermined unique display would in the dynamic display screens.

Even though the participants were asked to be accurate, a human error such as a misclick could still occur and this was taken into account. In the performance phase, participants performed each combination of the dependent variables four times (see table 10). One mistake such as a misclick for any given combination of the dependent variables was allowed and still considered accurate. Thus a participant was considered accurate for a combination of the dependent variables if 0 or 1 mistake was made.

The accuracy proportion was calculated by first summing up all accurately chosen displays for a given combination of the dependent variables by aggregating the data. This number was then divided by 4 to get to a number between 0 and 1.0. This meant that the accuracy score should be 0.75 or higher to be considered accurate for the given combination of the dependent variables. The accuracy proportion per person for each combination of the dependent variables can be found in tables in 13 for the static displays and in 18 for the dynamic displays. These tables allowed for visual identification of possible participants who either did not understand the tasks or who did not take the tasks seriously, and to identify possible problems with certain combinations of conditions.

The mean the accuracy score per display for each combination of dependent variables was calculated and presented in figure 13 and 18.

		static					
		Known state			Unknown state		
		8 displays	32 displays	72 displays	8 displays	32 displays	72 displays
		proportion accurate	proportion accurate	proportion accurate	proportion accurate	proportion accurate	proportion accurate
		Mean	Mean	Mean	Mean	Mean	Mean
display type	CCD	1.00	.97	1.00	.95	.95	.88
	ID4	.95	.95	.83	.78	.97	.92
	ID1	1.00	1.00	.98	.90	.90	.88

Figure 11: The accuracy score per combination of independent variables per display type for static displays.

		dynamic displays					
		Known state			Unknown state		
		8 displays	32 displays	72 displays	8 displays	32 displays	72 displays
		proportion accurate	proportion accurate	proportion accurate	proportion accurate	proportion accurate	proportion accurate
		Mean	Mean	Mean	Mean	Mean	Mean
display type	CCD	.95	.98	.95	.53	.85	.77
	ID4	.95	.90	.89	.63	.75	.86
	ID1	1.00	.98	.97	.43	.87	.88

Figure 12: The accuracy score per combination of independent variables per display type for dynamic displays.

5.3.2 Latency

The variable of main interest in this thesis was the efficiency of a display for different tasks. The efficiency was calculated by taking the median latency in seconds for each combination of independent variables (see section 5.2.1) corrected for accuracy. This meant that latency values from inaccurately identified displays were removed from the data. This was done because being accurate was the participants main task. Latency for static displays was the time measured in seconds from when the participant pressed the ctrl-key to uncover the grey boxes until the moment the participant pressed the spacebar. Latency for the dynamic displays was the time measured in seconds from the moment the anomaly display entered the anomaly state until the moment the participant pressed the spacebar.

Once the latency of inaccurately identified displays were removed, the median latency was calculated per person per combination of independent variables if its accuracy score was 0.75 or higher. This ensured that at least 3 or 4 data points were used to calculate the median latency per person per combination of independent variables. The data was aggregated using the variables *participant id, number of charts, chart type, dynamic and session type* as break variables. The aggregated variable was *latency accurate median*. The median was calculated since it is less influenced by outliers than the mean as a measure of central tendency.

5.4 Study set-up

All participants performed the experiments on the same PCs and screens. The participants were provided with a pre-test survey before they started the study on the computer and a post-test survey after they finished, to keep the current setup as similar as possible to the previous research. The full test layout of the test participants performed on screen can be found in appendix E.

5.4.1 Computers lab

The experiments where performed at Information Technology Centre of Campus Polacksbacken at Uppsala University in a computer lab which had 20 computers running Windows 10. The computers are distributed next to each other, with six computers along 2 walls and 2 rows of 4 in the center of the lab. All displays were HP Elite Display E272q 27-inch monitors, which had a native QuadHD (2560x1440) resolution and 60Hz refresh rate. Before participants entered, the displays were set to 100% contrast, 100% brightness and a neutral color temperature.

All participants took part in the data collection in the same room and on the same PCs and screens. Two data collection sessions were held on different days. Both data collection sessions were run in the afternoon. All participants began the evaluation at 14:15.

The number of PCs that were booted in the computer lab equaled the number of participants during the testing session plus one. The extra PC was booted up in case something would go wrong with one of the computers during the study. The testbed was loaded on each PC. The CCD was started on every third PC, then the ID4 was started on every PC next to the CCD. The ID1 was started on every PC next to the ID4. The pre-test survey (appendix C), post-test survey (appendix D) and a pen was placed on the desk in front of each PC.

On arrival, participants were given the consent form (appendix A) before entering the lab. The participants were randomly divided into thee groups when they entered the lab by selecting a PC themselves. There was no information on the start screen from which participants could deduct the display type they would be evaluating. The start screens on all PCs were equal except for the participant ID.

On screen, the participants were asked to first fill in the pre-test survey, before starting on the PC. After general information, participants performed the color pre-test again. Participants that did not pass the pre-test were unable to continue the evaluation. On screen, they were thanked for their participation and then would be gifted a movie ticket. Participants who passed the pre-test went on to the evaluation of the respective display. Upon completion of the evaluation, the participant was thanked on screen and asked to fill in the post-test survey. On the same screen they were told to hand-in the pre-test survey, post-test survey and consent form, if they had not yet handed it in, in order to collect their movie ticket.

5.5 Participants

The sampling technique used to recruit participants was a combination of convenience and snowball sampling. Participants were requited via channels such as Facebook groups, direct friends and classmates of the researcher. Even though non-probability sampling techniques were used, it was believed that a wide variety of demographics was achieved among the participants. However most participants were likely related to Uppsala university as they were students at the university.

The link that was distributed online led to a color pre-test. Participants were asked to perform this pre-test before signing up, to test whether they were able to distinguish between the chosen hues. Participants that scored less than 90% correct on the pre-test at home were discouraged from joining the study.

If the participants completed the pre-test with 90% correct they were considered ideal participants and invited to sign up for the evaluation. If the participants got less than 90% correct they were discouraged from signing up for the evaluation but not necessarily excluded. There was no limit as to how many times a potential participant could perform the color pre-test and no data from the test was collected in a database.

In total, 36 participants signed up for the study and 30 actually participated in the experiments. The participants were divided between two session days: Tuesday and Wednesday. The sessions started at the same time on both days: 14:15. For the Tuesday session 19 signed up and 17 signed up for the Wednesday session. This resulted in 17 participants who took part in Tuesdays study as 2 did not show up, and 13 participants took part in the Wednesday study as 4 did not show up.

5.5.1 Participant demographics

Lehane used a pre-test survey to understand the demographics of the participants [18]. The same pre-test survey was used in the current evaluation in order to replicate Lehanes study as much as possible. One question was removed, since it was considered controversial and no relevant conclusions could be drawn from the question in the previous thesis. The removed question asked participants which country they had lived in most of their life. Lehane added the question since some research suggested that Swedish students would perform differently to international students. This claim was not proven by Lehane. The pre-test survey used can be found in appendix C.

The demographics were explored using data from the pre-test survey. The 30 participants completed the survey, and no data was missing. This provided data on the participants gender, age, level of English and type of program studied. This data was analyzed to make a statement about the demographics and to check for potential confounding variables in the data. Since the participants distributed themselves between the display types, it was also important to check subsequently whether the distribution between the display types was equal.

The participants had a mean age of 26.40, with the maximum age being 39 and the minimum age 20. The gender of the participants was divided up to be 10 female participants and 20 male participants.

Out of the 30 participants, a majority of 22 had studied English as a subject in school for 11 or more years, 6 had studied English as a subject in school between 6 and 10 year, and 2 participants had studied English as a subject in school for 5 or fewer year.

Since a majority of the participants were university students (24 participants), they were asked what they studied. Out of 30, 10 studied an interdisciplinary discipline, 6 were pursuing a degree in social science, 8 were studying a STEM-related discipline, and 6 were not studying.

5.5.2 Division of participants between display types

The participants were equally and randomly divided into three groups based on the display type. A chi-square test of independence was done subsequently using the demographic data provided. This was done to check whether the spread of the participants was as equal as possible among the three display types.

No data was missing data for any of the variables. There was no significant relationship between gender and display type: X2 (2, N = 30) = 3.90, p = .23. Examining the relationship between display type and whether a participant was currently studying revealed no significance: X2 (2, N = 30) = 1.25, p = .85. The academic discipline did not reveal any significance either: X2 (6, N = 30) = 2.45, p = .96. Nor was there a significant relation between the years of English as a subject in school the participants had and the chart type: X2 (4, N = 30) = 6.18, p = .19. Finally, a one-way anova was used to test whether a relationship existed between the age of the participants and the display type they tested. No significant relationship was found F(2,26) = 1.242, p = .305.

5.5.3 Post-test survey

The post-test survey Lehane used asked participants, after completing the study, to rate their confidence of identifying the last four screens and what strategy they used [18]. This same post-test survey was used in the current study (D). The reason for asking participants questions regarding the last four screens was because these screens were the most recent screens the participants had completed and would still be in their memories. Furthermore these last four screens were regarded as the most difficult screens to complete, as they were both dynamic and showed 72 displays at the same time. The post-test survey could reveal vigilance issues.

5.5.4 Confidence

The first question on the post-test survey asked participants how confident they were that they had correctly identified the unique chart on the last four screens. Participants could answer the question by a five-point likert-scale [19]. This data was used to get an inside of the subjective perception of the participant on the display they tested. This perception was compared to the results of which display was most efficient.

This data was explored using a chi-square analysis to check whether any of the participants demographic data provided could be a confounding factor. The significance level used was $a = 0.05$. The data of one participant was excluded from the analysis since that particular participant did not start the 72-chart experiment.

There did not appear to be any relationship between confidence and whether the participant was studying (X2 (3, N = 29) = 5.26, p = .15), between confidence and gender (X2 (3, N = 29) = 2.17, p = .539), between the participants confidence and how many years of English as a subject in school they had (X2 (6, N = 29) = 10.62, p = .10), between the level of confidence and the academic discipline participants were studying (X2 (9, N = 29) = 12.82, p = .17), or between confidence and the display type the participant was assigned to (X2 (6, N = 30) = 6.067, p = .42). Whether a relationship existed between the participants age and their level of confidence was examined using a one-way ANOVA, but no relationship was found (F(2, 27) = 1.460, p = .250).

6 Experimental study

The aim of the experimental study was to support existing theories surrounding visual search and parallel processing. The ID was improved using color according to theories such as the pop-out effect. These theories have been implemented into two new designs: ID1 and ID4. The newly designed displays are predicted to perform similar or better than CCD.

6.1 Hypotheses

Based on the research aim, 6 hypotheses were posed. H1, H2 examined visual search in the displays. H3, H4, H5 and H6 examined how parallel processing is supported in the displays. First, similar hypotheses as for H1 and H2 were posed but for the static 32 displays where the unique state.

H1: ID1 allows for faster identification of a known unique state than CCD for the static 8 displays.

H2: CCD allows for faster identification of a known unique state than ID4 for the static 8 displays.

H3: ID1 allows for faster identification of a known unique state than CCD for the static 32 displays.

H4: CCD allows for faster identification of a known unique state than ID4 for the static 32 displays.

H5: The latency difference between 32 and 8 displays is smaller for ID1 than for CCD for the static displays where the unique state is known.

H6: The latency difference between 32 and 8 displays is smaller for CCD than for ID4 for the static displays where the unique state is known.

6.2 Methods

Established literature in visual search focuses on static displays, and its participants are told the unique feature they are looking for [13] [24] [27]. Therefore, only data from static displays in which the unique state was known was used in the experimental study. In addition was only data from 8-display screens and 32-display screens used. The accuracy and latency data collected in section 5 was used for this experiment. The data was prepared according to section 5.3.

While the main dependent variable of interest was latency, it was a requirement for the participants to be as accurate as possible. Appendix I contains a table with accuracy per participant for each combination of independent variables of static displays. This appendix was checked to see if any participants or individual data points needed to be excluded. No participants of the combination of independent variables static displays, known state were excluded from both the 8-display screens and 32-display screens. All participants of these combinations of independent variables had an accuracy proportion of 0.75 or higher. The mean accuracy proportion for each display was calculated using the median proportion per participant to see if there was a difference between displays or between the 32-display screens and 8-display screens. However some individual data points were excluded from analysis since they were inaccurate (see table 2).

display type	**8-displays**	**32-displays**
CCD	0	1
ID4	2	2
ID1	0	0

Table 2: Total number of removed data points for static displays with a known unique state per display type, for the 8-display screens and 32-display screens.

The median latency per person of each combination of independent variables was calculated (see appendix K). The median latency per display for each combination of independent variables was calculated using the median latency per person (see figure 14).

		number of displays	
		8	32
		proportion accurate	proportion accurate
		Mean	Mean
display type	CCD	1,00	,97
	ID4	,95	,95
	ID1	1,00	1,00

Figure 13: The mean accuracy proportion for each display for static displays with a known unique state.

		number of displays	
		8	32
		latency	latency
		Median	Median
display type	CCD	2,20	6,05
	ID4	1,92	4,48
	ID1	1,90	1,80

Figure 14: The mean latency in seconds corrected for accuracy per display for static displays with a known unique state.

6.3 Results

The differences in distributions of the latency data of the 8 static displays where the unique state was known to the participant; an interval variable, were examined between two groups. Figure 15 gives a rough idea how the distributions between the displays differ. The distributions of the latency data was explored. Results showed that the kutsosis was -0.606 and the skewness was 0.913 for ID4, kutrosis was -1.483 and the skewness was -0.072 for ID1, and kutrosis was -1.232 and the skewness was 0.323 for CCD. This showed that the data was not normally distributed. Thus a Mann-Whitney U test was performed on H1 and H2.

H1: ID1 allows for faster identification of a known unique state than CCD for the static 8 displays. A Mann-Whitney U test indicated that there was no significant difference ($N_{CCD} = 10$, $N_{ID1} = 10$, $U = 31$, $p = 0.083$) between the identification latency in seconds of ID1 (Mdn = 8.60) and CCD (Mdn = 12.40).

H2: CCD allows for faster identification of a known unique state than ID4 for the static 8 displays. A Mann-Whitney U test indicated that there was no

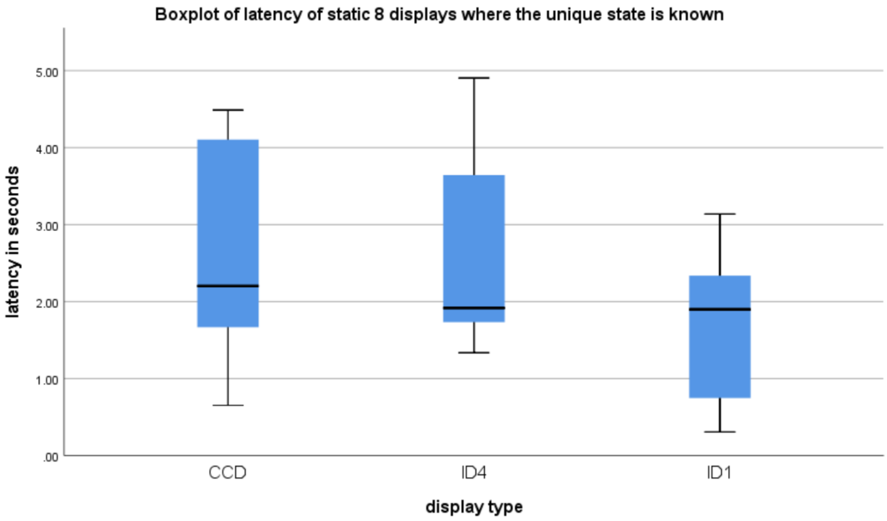


Figure 15: A boxplot of latency in seconds corrected for accuracy per display for the static 8 displays where the unique state is known.

significant difference ($N_{CCD} = 10$, $N_{ID4} = 10$, $U = 49$, $p = 0.485$) between the identification latency in seconds of CCD (Mdn = 10.60) and ID4 (Mdn = 10.40).

Hypotheses H3 and H4 compared the differences in distribution of the latency data of the 32 static displays where the unique state was known to the participant; an interval variable, between two groups. Figure 16 gives a rough idea how the distributions between the displays differ. The distributions of the latency data was explored. Results showed that the kutrosis was -1.197 and the skewness was 0.600 for ID4, kutrosis was 1.322 and the skewness was 1.364 for CCD, and kutrosis was -0.303 and the skewness was 1.012 for ID1. This showed that the data was not normally distributed. Thus a Mann-Whitney U test was performed on H3 and H4.

H3: ID1 allows for faster identification of a known unique state than CCD for the static 32 displays. A Mann-Whitney U test indicated that there was a significant difference ($N_{CCD} = 10$, $N_{ID1} = 10$, $U = 27$, $p = 0.0445$) between the identification latency in seconds of CCD (Mdn = 12.80) and ID1 (Mdn = 8,20).

H4: CCD allows for faster identification of a known unique state than ID4 for the static 32 displays. A Mann-Whitney U test indicated that there was no significant difference ($N_{CCD} = 10$, $N_{ID4} = 10$, $U = 48$, $p = 0.456$) between the identification latency in seconds of CCD (Mdn = 10.70) and ID4 (Mdn = 10.30).

Hypotheses H5 and H6 compared the differences in distribution of the latency difference between the 32 and 8 static displays where the unique state was known to the participant; an interval variable, between two groups. The line chart of figure 17 was created to get an initial understanding of the extent of the difference in latency between the display types. The line chart suggests to

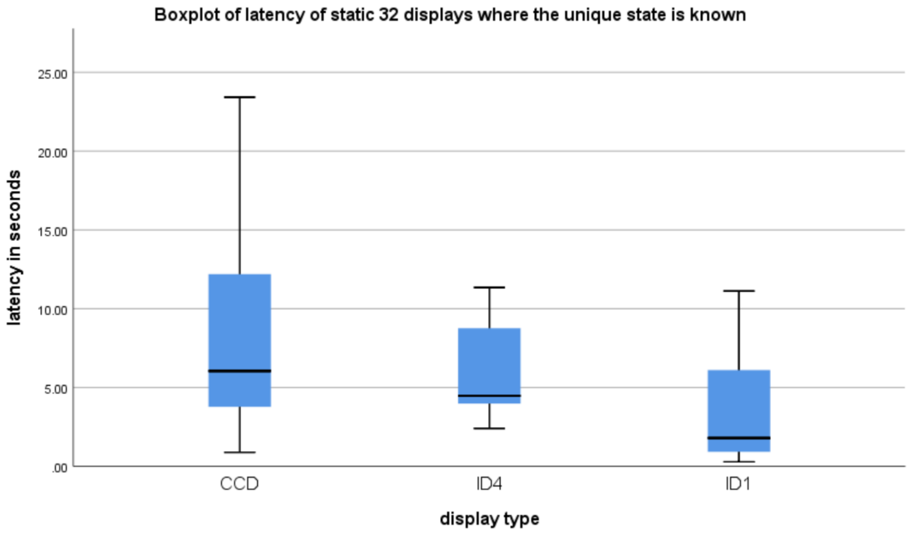


Figure 16: A boxplot of latency in seconds corrected for accuracy per display for the static 32 displays where the unique state is known.

confirm the prediction that ID1 is the most efficient display type.

First the difference was calculated by subtracting the latency in seconds of the 32 display with the latency in seconds of the 8 display per participant per combination of conditions (see appendix M). The distributions of the latency differences was explored per display type. Results showed that the kutrosis was 3.066 and the skewness was 1.626 for ID4, kutrosis was -0.112 and the skewness was 1.114 for ID1, and kutrosis was 2.149 and the skewness was 1.566 for CCD. This showed that the data was not normally distributed. Thus a Mann-Whitney U test was performed on H5 and H6.

H5: The median latency difference in seconds between 32 and 8 displays is smaller for ID1 than for CCD for the static displays where the unique state is known. A Mann-Whitney U test indicated that there was a significant difference ($N_{CCD} = 10$, $N_{ID1} = 10$, $U = 24$, $p = 0.026$) between the median latency difference in second of CCD (Mdn = 13.10) and ID1 (Mdn = 7.90).

H6: The median latency difference in seconds between 32 and 8 displays are smaller for CCD than for ID4 for the static displays where the unique state is known. A Mann-Whitney U test indicated that there was no significant difference ($N_{CCD} = 10$, $N_{ID4} = 10$, $U = 47$, $p = 0.4265$) between the median latency difference in second of CCD (Mdn = 10.80) and ID4 (Mdn = 10.20).

6.4 Experimental study conclusion

H1, H2, H3 and H4 examined which display type supported visual search the best. The results suggest that adding color to the ID display allows for similar

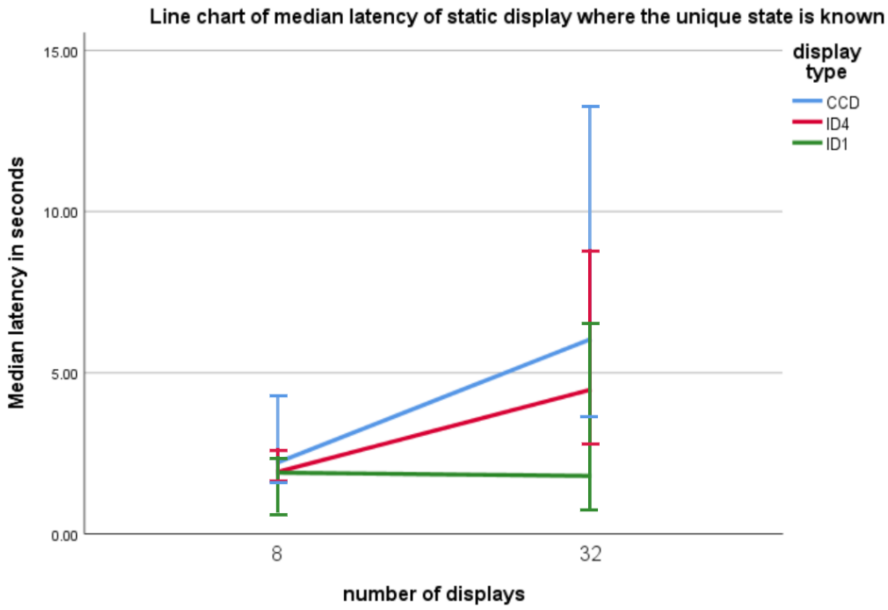


Figure 17: A line chart with median latency in seconds per display with a known unique state for the static 32 and 8 display screen.

or better efficiency than the CCD display.

H1 and H3 proved that the ID1 is significantly more efficient than the CCD display. These results suggest that the addition of color to ID allows it to become more efficient than the CCD.

H2 and H4 revealed that there is no significant difference in efficiency between the CCD and ID4. This means that the addition of color to ID alone does not make the display more efficient. How color is added has more influence of the efficiency as proven by H1 and H3.

H5 and H6 investigated the efficiency difference between the different display types to see which display type supports parallel processing the best. The lower the latency difference of a display the better is supports parallel processing. Thus the display with the lowest latency difference supports parallel processing the best. The results showed that the addition of color to the ID display allows for similar or better parallel processing compared to the CCD display. The significant difference found in H5 suggests that the addition of color to the ID allows for better parallel processing. H6 revealed no significant difference which indicates that the addition of color to ID allowed for a similar extent of parallel processing.

The fact that ID1 is more effective than CCD when it comes to visual search can be explained using the pop-out effect. The dominant state in ID1 is the only feature where color is applied. The feature difference between the dominant state and the rest of the display is big enough to enhance the saliency of the dominant state so that it pops out. Users of ID4 can simply apply bottom-processing where as users of CCD have to apply top-down processing, a cognitive-intensive process.

The fact that parallel processing is supported better by ID1 than CCD can also be explained using the pop-out effect. A pop-out effect also occurs for the display in an anomaly state among the other displays. Even-though colors of the states have the same brightness level, they are different enough to stand apart from each other. Since each state has a unique color, they can quickly be recognized. The ID1 allows for parallel search efficiently since all displays in the set can be attended to all at once regardless if the set size is 8 or 32.

7 Exploratory study

The exploratory study aimed to explore the data in a way that is relevant for more general settings. Ecological validity was of main relevance. The exploratory study aimed to connect to the results from the experimental study (see section 6). Data in more general settings is unpredictable; it is often unknown what state would be the anomaly state. The real-world is dynamic and thus operators need to deal with changing, dynamic data. Thus present data from the dynamic displays and unknown states ecological validity. The exploratory study aimed to explore visual search and parallel processing in these data.

Due to the ecological validity of the exploratory study and due to the lack of related literature it becomes more difficult to predict what will happen. Therefor were these data explored without hypotheses. These data was however explored similarly to the experimental study. The extent to which visual search and parallel processing is supported was explored. The results of this study provided hypothesis that could be tested in future studies on dynamic displays with an unknown unique state.

7.1 Methods

The exploratory study aimed to cater to the interest in more general settings. This experiment is applicable to real life as it deals with dynamic content and unknown states. The results of dynamic displays with an unknown unique states were analysed for the 8, 32 and 72 displays. Since little literature has previously investigated visual search and parallel processing in dynamic displays and because of the high ecological validity the results are unpredictable. Therefor no hypothesis were formed for the analysis. All displays that were dynamic and had an unknown unique state were analysed because the exploratory study aimed for ecological validity. The data was prepared according to section 5.3.

The exploratory study was set up according to methods presented in section 5. Participants performed tasks with dynamic displays in which the unique state was unknown to the participant. Participants received no special training's phase related to dynamic displays with an unknown unique state as the training received before the experimental study was deemed enough. The task only had a performance phase.

Before stating the exploratory study participants received instructions on the task and were provided with an example. In the instructions the participants were not told which state would be unique. At the beginning of the task, no display was in the unique unknown state. All displays would dynamically move into different states until one display would enter a state different to the other displays. It was the participant‘s task to locate and select the display that was in a state different from the other displays as quickly and accurately as possible after this display entered the unique state. Participants performed this task for screens that showed 8, 32 an 72 displays. For each number of displays shown on screen they repeated the task 4 times.

Latency in seconds was the dependent variable of main interest. However, it was a requirement for the participants to be as accurate as possible. First, the accuracy proportion was calculated per person (see appendix J). Then the mean accuracy proportion was calculated using the accuracy proportion was calculated per person to see if there were any problems with particular combinations of independent variables (see figure 18). According to figure 18 is the mean accuracy proportions for the dynamic displays with an unknown unique state less than 0,75 in the 8-display screen. Figure 18 shows that participants of each display type were particularly inaccurate for the 8-display screens. This suggests that there might have been a problem with this combination of independent variables. Thus was the data from the dynamic displays of 8-display screens with an unknown unique state excluded. Data from the remaining combinations of independent variables included the 32-display and 72-display screens. The number of individual data points that were excluded from the data is displayed in table 3.

		number of displays		
		8	32	72
		proportion accurate	proportion accurate	proportion accurate
		Mean	Mean	Mean
display type	CCD	,53	,85	,77
	ID4	,63	,75	,86
	ID1	,43	,87	,88

Figure 18: The mean accuracy proportion per display per combination of independent variables for dynamic displays.

display type	8-displays	32-displays	72-displays
CCD	40	8	9
ID4	40	11	11
ID1	40	5	5

Table 3: Total number of removed data points for dynamic displays with an unknown unique state per display type, for the 8-display screens and 32-display screens.

The mean accuracy proportion is less than 1.0 for the 32-display and 72-display screens (see figure 18). This means that it is likely that the accuracy proportion of some participants was lower than 0.75. Appendix J shows that 5 participants had an accuracy proportion of less than 0.75 in the 32-display screens. These included 2 participants of the CCD, 2 participants of the ID4, and 1 participant of the ID1 display. Of the 72-display screens, there were 5 participants with an accuracy proportion of less than 0.75. Of which 2 participants tested the CCD, 2 participants tested the ID4, and 1 participant tested the ID1 display. These 10 cases were removed from the data.

The median latency per person for each combination of independent variables was calculated using the accurate guesses only. Per person for each combination of independent variables there were 3 or 4 data points to be used for analysis, since participants with an accuracy proportion with less than 0.75 were removed from the data. The median latency per person for each combination of independent variables was calculated (appendix L). The median latency per display was calculated for the 32-display screens and the 72-display screens (figure 19.

		dynamic displays					
		Known state			Unknown state		
		8 displays	32 displays	72 displays	8 displays	32 displays	72 displays
		median latency	median latency	median latency	median latency	median latency	median latency
display type	CCD	3.63	10.17	20.41	10.27	9.84	22.51
	ID4	4.47	6.36	8.11	6.95	7.14	10.22
	ID1	3.48	4.37	3.87	4.01	3.51	3.88

Figure 19: The mean latency in seconds corrected for accuracy per display per combination of independent variables for dynamic displays.

7.2 Results

The exploratory study aimed to connect its results to the experimental study (see section 6). The data was analysed similarly to the experimental study in section 6. This meant that the analysis investigated visual search and parallel processing with alpha = 0.5. Additionally, fatigue in the 72-display screens was explored.

7.2.1 Visual search

The exploratory study analysed which display type supports visual search the best. A Mann-Withney U test was performed on the data since this was also done in the experimental study.

First the 32-display screens were analysed. The ID4 and ID1 were individually compared against the CCD. Both comparisons revealed a significant difference. A Mann-Whitney U test indicated that there was a significant difference ($N_{CCD} = 8$, $N_{ID1} = 9$, $U = 5.000$, $p = 0.003$) between the identification latency in seconds of ID1 (Mdn = 12.88) and CCD (Mdn = 5.56). A Mann-Whitney U test indicated that there was a significant difference ($N_{CCD} = 8$, $N_{ID4} = 8$, $U = 11.000$, $p = 0.014$) between the identification latency in seconds of CCD (Mdn = 11.13) and ID4 (Mdn = 5.88).

The ID4 and ID1 were compared against the CCD for the 72-display screens. Here too, significant differences were found. A Mann-Whitney U test indicated that there was a significant difference ($N_{CCD} = 8$, $N_{ID1} = 9$, $U = 0.000$, $p = 0.000$) between the identification latency in seconds of ID1 (Mdn = 5.00) and CCD (Mdn = 13.50). A Mann-Whitney U test indicated that there was a significant difference ($N_{CCD} = 8$, $N_{ID4} = 9$, $U = 0.000$, $p = 0.003$) between the identification latency in seconds of CCD (Mdn = 12.75) and ID4 (Mdn = 5.67).

7.2.2 Parallel processing

The data in the exploratory study was also explored for differences in the extent to which the different display support parallel processing. Only data from participants with an accuracy proportion of 0.75 or higher on both the 32-display and 72-display screens were considered. The data from 4 participants were removed before analysis.

First, the latency difference in seconds was calculated by subtracting the latency in seconds from the 72-display screens from the 32-display screens per person (see appendix L). Using these results, the median latency difference in seconds was calculated per display. These initial results suggest that parallel processing is best supported in ID1, followed by ID4, and finally CCD in third place.

An overview was created by plotting the latency data of the 32 and 72 displays in a line chart in figure 20. This line chart suggests to confirm the posed prediction that ID4 is more efficient than CCD and that ID1 is the most efficient display type. It also shows that parallel processing is likely supported best by ID1. However, it also suggests that ID4 might also support parallel processing well, or at least much better than CCD.

The median latency difference was compared between ID1 and CCD, and between ID4 and CCD. While a significant difference was found for the comparison of ID1 and CCD, no significant difference was found when comparing CCD to ID4. A Mann-Whitney U test indicated that there was a significant difference (N_{CCD} = 8, N_{ID1} = 9, U = 15.000, p = 0.023) between the difference in latency in seconds between the 72 and 32 display screens of ID1 (Mdn = 6.67) and CCD (Mdn = 11.63). A Mann-Whitney U test indicated that there was no significant difference (N_{CCD} = 8, N_{ID4} = 8, U = 15.000, p = 0.141) between the median difference in latency in seconds between the 72-display and 32-display screens of CCD (Mdn = 10.13) and ID4 (Mdn = 6.88).

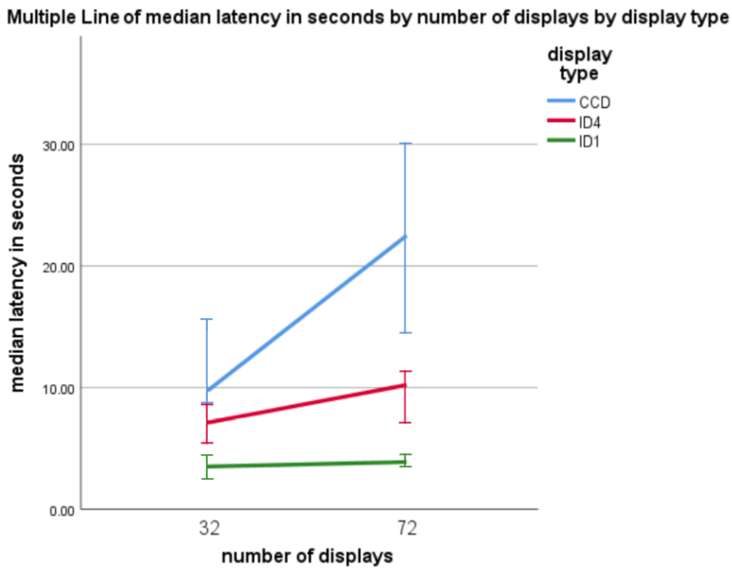


Figure 20: A line chart with median latency in seconds per display with an unknown unique state for the dynamic 32 and 72 displays.

7.2.3 Fatigue in 72-display screens

Previous research expressed concern regarding fatigue in participants [18] [20]. Especially the dynamic 72-display screens in which the unique state was unknown could be prone to fatigue as they were the last combination of independent variables the participants tested. The accuracy of the combination of independent variables was checked since low accuracy in the last displays could be a sign of fatigue. The data from the post-test survey participants filled in after they had completed the study on screen was also used to explore perceived fatigue.

Figure 18 suggests no dramatic difference in mean accuracy per display between the 32- and 72- display screens. The ID1 and ID4 seem to even be slightly more accurate for the 72-display screens than for the 32- display screens. These differences do not seem to indicate fatigue during the last combination of independent variables.

The data from the two questions in the post-test survey asked were analyzed to explore different strategies participants had to identify the correct display and how confident they felt that they had identified the correct display.

The post-test survey asked two questions: 1. How confident are you that you identified the unique chart on the last four screens? 2. What was your way of finding the unique chart on the last four screens? The second questions was analyzed using thematic analysis. The first step was to remove the participant who did not perform the 72 display experiment from the analysis as the question dealt with the last four screens of the 72 display experiment. Then the answers were sorted into three piles, each representing a condition (display type). Themes were derived by going through each pile to find common strategies. These strategies were labeled as the themes. During this first round of thematic analysis, six themes were found. These themes are described bellow, including the number of participants who reported the strategy per display type.

1. Missing state: Participant notice there is one state missing. They look for that state.
   - CCD: 6
   - ID1: 3
   - ID4: 8
2. Nothing specific: Participants report not having a particular approach.
   - CCD : 2
   - ID1: 1
3. State-by-state: Comparing the state of a single chart to others.
   - CCD : 1
   - ID4: 1
4. Learning: Participants report that it was easier to spot since they had been doing it over and over.
   - ID1: 1
5. Overview: Participants sit back to watch all charts at the same time.
   - ID1: 4
6. Unknown: Participants reported strategies that were not understandable.
   - CCD : 1
   - ID1: 1

Theme 1, missing state, was bigger than the other themes. In total it contained 17 participants. Therefore it was broken down further by checking if another theme could be discovered within the theme.

1. Comparing states: Comparing a state from one chart to another so check if it reoccurs.
   - ID4: 2

2. Missing/least frequent state: looking straight for the missing state or the one that occurs least, that is the unique one.
   - CCD : 3
   - ID1: 1
   - ID4: 2
3. Movement: observing which state was avoided, that is the unique state.
   - CCD : 2
4. Overview: Participants sat back to look at all charts at the same time.
   - ID1: 1
5. Eliminating states: Finding states were present, eliminating those as the possible unique state.
   - CCD : 1
   - ID1: 1
   - ID4: 3
6. Color: Identifying which color were predominantly present, then concluding the missing color is the unique state.
   - ID4: 1

7.2.4 Confidence

The first question on the post-test survey asked participants how confident they were that they had correctly identified the unique chart on the last four screens. Participants could answer the question by a five-point likert-scale [19].

The relationship of interest was examined using chi squared analysis between the condition (display type) the participants were assigned to and their reported level of confidence. No significant relationship could be found between display type and the reported level of confidence (X2 (6, N = 29) = 6.07, p = .42).

7.3 Exploratory study conclusion

The exploratory study examined dynamic displays with a dominant state unknown to the observers. These type of displays, while not often studied in literature on visual search or parallel processing is interesting for general situations. The experiment with dynamic displays with a dominant state unknown to the observer, closely resembles unanticipated and unfamiliar events in a dynamic environment such as the real world. The screens that displayed 32 and 72 dynamic displays with an unknown dominant state were examined. The results from the screens with 8 displays were discard due to high inaccuracy.

Similar to the experimental study, visual search seems faster for ID1 than for CCD for both the 32 and 72 display condition. Unlike in the experimental study, it seemed that visual search was faster for the ID4 than the CCD for both the 32- and 72- display screens. These results could possibly be explained by the three search tasks: parallel search, serial search and guided search. From the results in both the experimental study and the exploratory study it seems that a search using ID1 is a parallel search task. Searches using CCD seems to be a serial search task from the results in the experimental study and the exploratory study. Based on the fact that ID4 allowed for faster visual search than CCD in the exploratory study while not in the experimental study could suggest that this display allows for a guided search task. The strategies ID4 participants used also hint towards it being a guided search task. Both a missing/least frequent state strategy and a comparing states strategy were reported by the participants of ID4. In a guided search task a subset of possible target items is filtered out based on basic visual features [43]. It might be that ID4 allows observers to create subsets of displays in a specific state based on which colors is most or least prominent on screen. That ID1 was the most efficient display is inline with the results from the experimental study. The latency and post-test questionnaire results suggest that the pop-out effect is also present when the anomaly state is unknown and displays are dynamic. A prominent strategy only ID1 participants reported using was overview which does not seem to be a cognitive-intensive strategy. Participants observed all displays at the same time until one entered the anomaly state.

The latency difference in seconds between the 32-display screens and the 72-display screens were used to analyse parallel processing. From the analysis it seems that parallel processing is better supported in ID1 than CCD. No significant difference found between CCD and ID4. These results are in line with the results from the experimental study. Even though the experimental study and related previous research focus on static displays with an unknown anomaly, it could be that similar effects happen in dynamic displays with an unknown anomaly. Therefore could the results of the exploratory study also be explained using the pop-out effect and previous research on parallel searching. The results from the post-test questionnaire backs this claim. Only ID1 participants used the overview technique to find the display with a unique state in the 72-display screen. Rather than scanning individual displays they sat back to watch all displays on screen at the same time.

8 Conclusion & Discussion

The aim of this thesis was to build upon previous research and development of the ID display [18] [20] and to improve its efficiency over the CCD display. Both visual search of a unique state in different quantities of displays displayed and parallel processing of the ID display were improved. Based on the aim posed:

> *Improve the efficiency of the ID using color so that the ID becomes more efficient than the CCD.*

The use of color in the ID aimed to improve the efficiency and effectiveness of the ID display. From theory four colors were derived. Using these same colors two new versions of the ID were created: the ID4 and the ID1. The newly design ID1 and ID4 were tested against the current industry standard: CCD. Two experiments were conducted in which the CCD, ID4 and ID1 were evaluated using the same tasks as where used in Lehanes thesis [18]. The tasks required participants to spot the display that was in a unique state compared to the other displays shown on screen. This task was performed by showing 8, 32 or 72 displays on screen.

The experimental study (see section 6) aimed to connect the research question to existing theories on visual search and parallel processing. For this experiment static displays were used. Participants knew which state was unique before hand and tested both 8 or 32 displays on screen.

The exploratory study (see section 7) explored visual search and parallel processing in displays with high ecological validity. Therefore was the unique state unknown to the participants and were the displays dynamic. Participants tested screens that showed 8, 32 and 72 displays. Since analysis showed that screens with 8 display had a low accuracy score, the results were not analyzed.

The research question was answered by looking at research on visual search and parallel processing and the results from both experiments. The ID1 display (see figure: 8) provides the most suitable answer to the research question. The use of the popup effect was most efficient in static displays with a known unique state and dynamic displays with an unknown unique state. The ID4 display (see figure: 9) could also be seen as a suitable solution, but only when displays are dynamic and have an unknown state. Both ID improvements use colors that all have the same perceived brightness and are placed strategically on the parts of the display that convey the information of relevance. However, the ID4 display proved that the placement of the colors and the colors itself were not enough to improve the efficiency of ID in both static and dynamic displays. The application of the pop-out effect to ID1 improved the ID so that it became more efficient than the CCD.

8.1 Visual search

The results from the experimental and exploratory study suggests that visual search in static displays is more efficient in ID1 than CCD. From the results found in this thesis it seems that visual search in dynamic displays with an unknown unique state follow the literature on static displays.

Interestingly, results of the exploratory study suggested that ID4 was more efficient for visual search than CCD in dynamic displays with an unknown state. While the ID1 and ID4 were not compared with each other, it is likely, according to theory on the pop-out effect [36], that ID1 would be more efficient than ID4. Its also not unlikely to think that the pop-out effect of ID1 aids more in static environments than dynamic ones. The ID1 only shows a fully saturated color when one state is dominant, while ID4 always shows all states in fully saturated colors. So the users task in ID1 is to judge the most saturated color, while the users task in ID4 is to judge the size of a states area. The color in ID4 is there to aid the user in that task.

The results of the exploratory study that suggested that ID4 was more efficient for visual search than CCD in dynamic displays with an unknown state was not expected or achieved in the experimental study. It could be that ID4 allows for guided search only when displays are dynamic, the unique state is unknown or a combination of these two factors. In that case, ID4 allows observers for filtering out a subset of possible target items according to the color and size of the dominant state when displays are dynamic and the unique state is unknown [43]. The strategies participants reported in the post-test survey also hinted towards ID4 allowing for guided search. However, it should be noted that in dynamic setting always one state was missing on screen load. This could have lead participants to search for the state and corresponding color that was missing on screen load, and then focus on the state or color that was missing. The results from the post-test survey suggest that participants of the ID4 used such a strategy. Out of 10 participants, 8 reported that they noticed that there was a missing state and searched for that state to find the anomaly state. However, this doe not mean that the ID4 might not support guided search. In fact, it would be further prove that the ID4 supports guided search if participants were able to filter out the colors that were present on screen load.

8.2 Parallel processing

As expected from literature on parallel processing, results from the experimental study showed that ID1 supported parallel search better than CCD in static display with a known unique state. Similar results were found in the exploratory study for dynamic displays with an unknown unique state. These results could be explained with the pop-out effect, similarly as the results on visual search. Especially since there no difference between CCD and ID4 in both experiments. It could be carefully said that parallel processing of dynamic displays with an unknown unique state follow the same basis in theory as static displays with a known unique state. Since similar results as in literature on parallel processing in static displays were found for dynamic unknown content.

8.3 Vigilance

Unlike previous research no problem with fatigue seemed to exist in the exploratory study [18] [20]. Thus vigilance was likely not an issue. The accuracy difference between 72- and 32-display screens were too small to cause problems. A similar trend was found when analysing the post-test survey as most participants where fairly confident that they found the correct display in the last four screens.

8.4 Previous work

This thesis aimed to improved the ID display designed by Laaksoharju and Lind, and studied by Löfvenberg [20] and Lehane [18]. While Lehane proved that the ID display performed only slightly better than CCD, this thesis proved that the ID with color added supports visual search better than the CCD.

A formal analysis cannot be performed to compare the current results with those collected by Lehane [18] because both studies did not have a random sample. Instead, a visual, qualitative comparison was done by creating box-plots of both data. The data from previous work was restructured in the same way as the data collected in this thesis to create a fair comparison. The common condition in this thesis and Lehanes thesis is the CCD display. Both studies have found similar results if the data is similar enough when visually comparing.

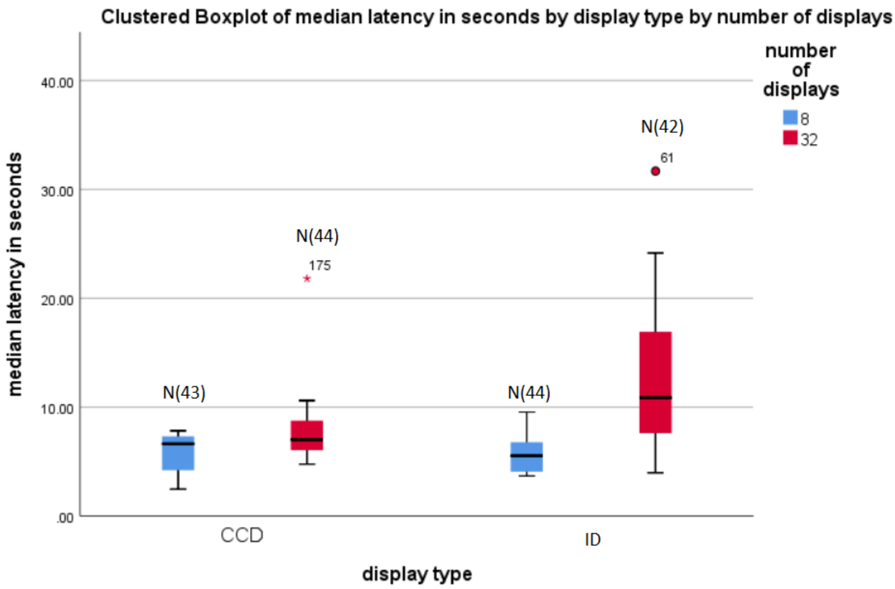


Figure 21: A boxplot of the 8 and 32 static displays with a known unique state. The data is taken from Lehanes work [18]. The data is from this thesis. The number of samples used per display for each number of displays is noted above each boxplot.

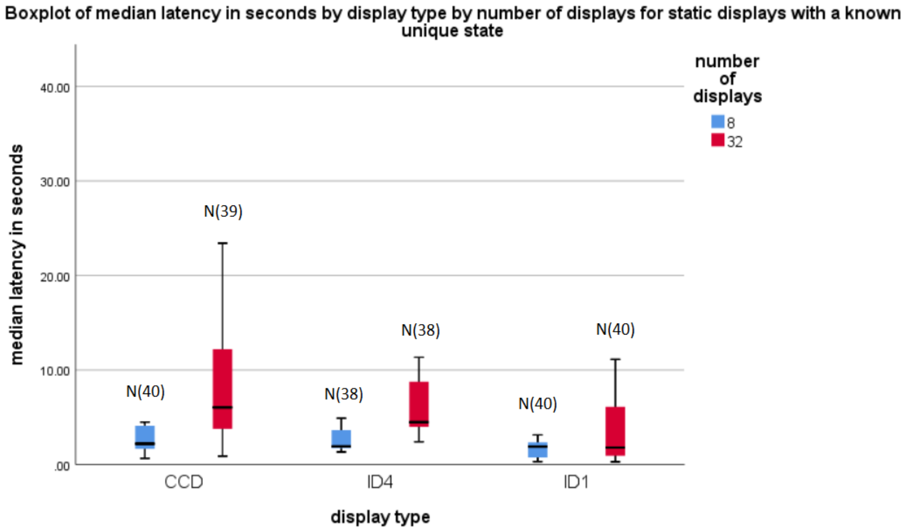


Figure 22: A boxplot of the 8 and 32 static displays with a known unique state. The data is from this thesis. The number of samples used per display for each number of displays is noted above each boxplot.

A visual comparison of the data from the CCD between figure 24 and 23 suggests that the results are different for the static displays with an known unique state between this thesis and Lehanes thesis. While Lehanes data shows a left-skewed distribution, this thesiss data show a right-skewed distribution when 8 displays are shown on screen. When 32 displays are shown on screen it seems that Lehanes data is more spread out than this thesiss data. Any conclusions made using these box-plots should be taken with high caution.

Comparing the ID with the ID1 and ID4 when 8 displays are shown on screen suggests that the ID1 and ID4 out-performs the ID. Comparing the results of when 32 displays are shown on screen suggest that the ID1 and ID4 are more efficient than the ID. Especially the spread of the distribution of the ID4 is smaller than the ID. This means that it is likely that the ID4 performs more consistently than the ID. The comparison, overall, hints that visual search is supported better by the ID1 and ID4 than by the ID studied by Lehane.

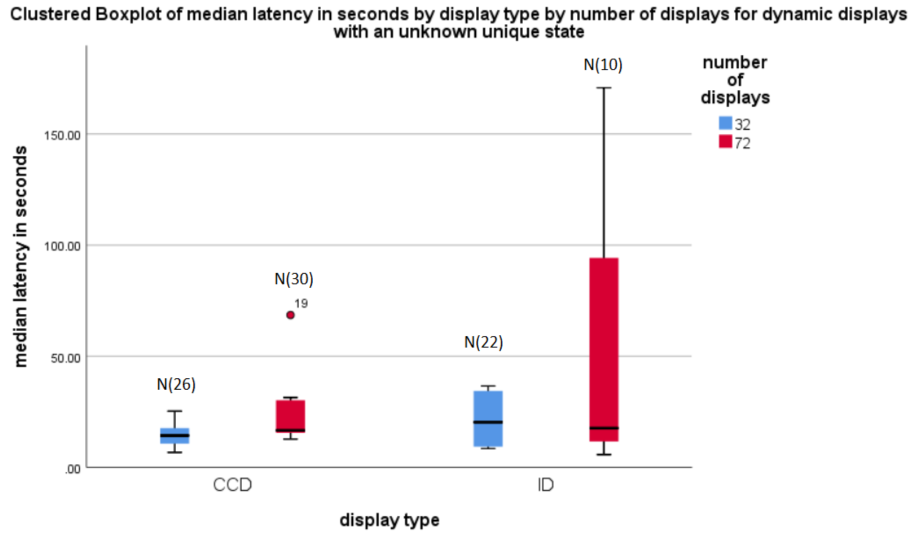


Figure 23: A boxplot of the 32 and 72 dynamic displays with an unknown unique state. The data is taken from Lehanes work [18]. The data is from this thesis. The number of samples used per display for each number of displays is noted above each boxplot.

The results from the data sets are fairly similar for the dynamic displays with an unknown unique state when visually comparing the data from the CCD display in figure 24 and 23. Visually comparing the results for the ID display in Lehane study with the results for the ID1 and ID4 in this study reveals that the improved ID displays were more efficient when it comes to visual search. Besides, the spread of the data from this thesis was not as spread out as Lehanes data. The efficiency of the participants of the ID4 and ID1 was more often centred around the median compared to the ID. This means that it is likely that the ID4 and ID1 performs more consistently than the ID.

These conclusions should be taken with caution, however. The participants in Lehanes study performed highly inaccurate for the dynamic displays with an unknown condition; meaning that they had an accuracy proportion of less than 0,75. Out of 32 participants, 16 were removed from the data for the 32 display screens. For the 72 display screens, 16 out of 31 participants were removed from the data, and thus not included in the boxplot shown in figure 23.

8.5 Limitations

A limitation of this thesis is that effect sizes were not be calculated for the differences for Mann-Withney U tests. Even-though significant differences were found, effect sizes could tell more about practical significance. A big effect size for differences between ID1 and CCD would be more evidence that the ID1 should be considered in a more general setting.

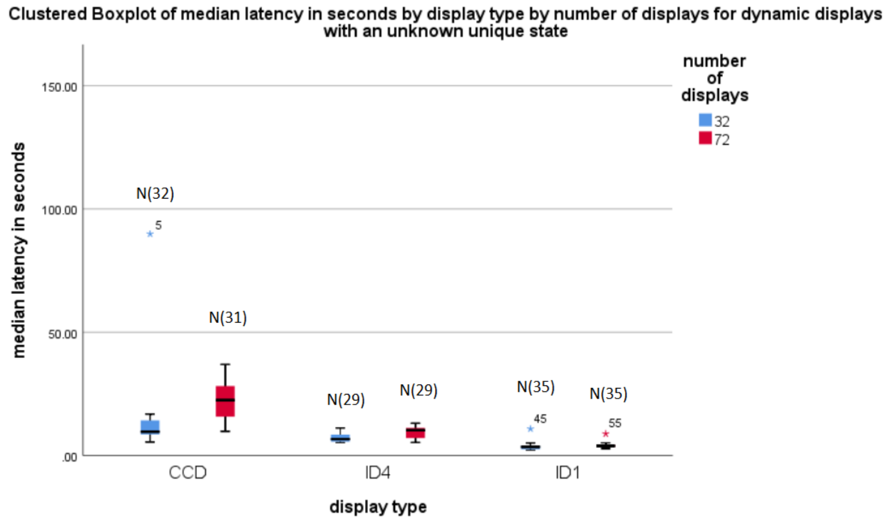


Figure 24: A boxplot of the 32 and 72 dynamic displays with an unknown unique state. The data is from this thesis. The data is from this thesis. The number of samples used per display for each number of displays is noted above each boxplot.

Another limitation was the order in which the participants tested the different combinations of independent variables between the experimental study and the exploratory study (see figure 25). This makes it difficult to compare the results from the exploratory study with the experimental study. In the experimental study, there were 4 screens between the static 8-display screens and the static 32-display screens. In the exploratory study, there were 16 screens between the dynamic 32-display screens and the 72-display screens. A learning effect could have influenced the participants results in the exploratory study.

Task	No. Displays	Dynamicity	No. Screens
Sequencial Display Training	8	Static	1
		Dynamic	1
	32	Static	1
		Dynamic	1
Training - Known Unique State	8	Static	4
		Dynamic	4
	32	Static	4
		Dynamic	4
Performance - Known Unique State	8	Static	4
		Dynamic	4
	32	Static	4
		Dynamic	4
	8	Static	4
		Dynamic	4
Performance - Unknown Unique State	32	Static	4
		Dynamic	4
Training - Known Unique State	72	Static	4
		Dynamic	4
Performance - Known Unique State	72	Static	4
		Dynamic	4
Performance - Unknown Unique State	72	Static	4
		Dynamic	4

Figure 25: The order in which the participants tested the conditions in the experiments. The data used in the exploratory study is marked by a red box. The data used in the exploratory study is marked by the blue boxes.

8.6 Future work

First should the difference in results of the ID4 between the experimental and exploratory study be further investigated. The limitations mentions that a learning effect could exist in this thesis since the order in the exploratory study was different than the order in the experimental study. It is advised to investigate whether a learning effect occurred in the data of this thesis.

On the other hand, more research is needed into visual search and parallel processing in dynamic content with an unknown unique state. It would be interesting to know if the results of the exploratory study will hold for the ID4 and how its efficiency compares to CCD and ID1. Based on the outcome of such a study, another study could investigate why the ID4 is more efficient than the CCD when the displays are dynamic and the unique state is unknown and why those results are not achieved when the displays are static and the unique state is known. The results found in the exploratory study are a suitable starting ground for such a further research. A future study could investigate visual search an parallel processing using dynamic displays in which the unique state is unknown to the participants by replicating the exploratory study. Following hypotheses posed in the discussion of the exploratory study can be used to further investigate visual search an parallel processing using dynamic displays in which the unique state.

> Visual search H1: *ID1 allows for faster identification of an unknown unique state than CCD in dynamic displays.*

> Visual search H2: *ID4 allows for faster identification of an unknown unique state than CCD in dynamic displays.*

> Parallel processing H1: *The latency difference between 72- and 32-display screens is smaller for ID1 than for CCD for the dynamic displays in which the unique state is unknown.*

References

[1] Alexey Anokhin, Alexey Ivkin, and Sergey Dorokhovich. Application of ecological interface design in nuclear power plant (npp) operator support system. *Nuclear Engineering and Technology*, 50(4):619–626, 2018.

[2] Lisanne Bainbridge. Ironies of automation. pages 775–779. Pergamon Press Ltd., 1983.

[3] Kevin B Bennett and John Flach. Ecological interface design: Thirty-plus years of refinement, progress, and potential. *Human factors*, 61(4):513–525, 2019.

[4] Kevin B Bennett and John M Flach. Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human factors*, 34(5):513–533, 1992.

[5] Kevin B Bennett and John M Flach. *Display and interface design: Subtle science, exact art.* CRC Press, 2011.

[6] Kevin B Bennett, Allen L Nagy, and John M Flach. Visual displays. *Handbook of human factors and ergonomics*, pages 659–696, 1997.

[7] KS Bibby, F Margulies, JE Rijnsdorp, RMJ Withers, and IM Makarov. Man's role in control systems. *IFAC Proceedings Volumes*, 8(1):664–683, 1975.

[8] Clark Borst, John M Flach, and Joost Ellerbroek. Beyond ecological interface design: Lessons from concerns and misconceptions. *IEEE Transactions on Human-Machine Systems*, 45(2):164–175, 2015.

[9] Robert Bridger. *Introduction to ergonomics.* Crc Press, 2008.

[10] Gershon Buchsbaum and Allan Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal society of London. Series B. Biological sciences*, 220(1218):89–113, 1983.

[11] Ronald Comer, Elizabeth Gould, and Adrian Furnham. *Psychology.* Wiley, 2013.

[12] Nick Dinadis and Kim J Vicente. Ecological interface design for a power plant feedwater subsystem. *IEEE Transactions on nuclear science*, 43(1):266–277, 1996.

[13] John Duncan and Glyn W Humphreys. Visual search and stimulus similarity. *Psychological review*, 96(3):433, 1989.

[14] John M Flach, Fumiya Tanabe, Kazuo Monta, Kim J Vicente, and Jens Rasmussen. An ecological approach to interface design. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 295–299. SAGE Publications Sage CA: Los Angeles, CA, 1998.

[15] Richard L Henneman and William B Rouse. Human problem solving in dynamic environments. understanding and supporting operators in large-scale, complex systems. Technical report, Georgia institution of technology, Center for man-machine systems research, 1987.

[16] Jerred Holt, Kevin B Bennett, and John M Flach. Emergent features and perceptual objects: re-examining fundamental principles in analogical display design. *Ergonomics*, 58(12):1960–1973, 2015.

[17] Shinobu Ishihara. *Tests for colour-blindness*. Kanehara Shuppan Company Japan, 1975.

[18] Darren Lehane. Visualising the unpredictable: Effective communication of process anomalies requiring human intervention, 2018.

[19] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[20] Johanna Löfvenberg. Poietic design: Heuristics and applications, 2016.

[21] Norman H Mackworth et al. Researches on the measurement of human performance. *Researches on the Measurement of Human Performance.*, (268), 1950.

[22] Norio Naito, Juichiro Itoh, Kazuo Monta, and Maomi Makino. An intelligent human-machine system based on an ecological interface design concept. *Nuclear engineering and design*, 154(2):97–108, 1995.

[23] Ken Nakayama, Gerald H Silverman, et al. Serial and parallel processing of visual feature conjunctions. *Nature*, 320(6059):264–265, 1986.

[24] Ulric Neisser. Visual search. *Scientific American*, 210(6):94–103, 1964.

[25] Harold Pashler. *Encyclopedia of the Mind*, volume 1. SAGE, 2013.

[26] Charles Poynton. The magnitude of nonconstant luminance errors. *A Technical Introduction to Digital Video*, 1997.

[27] Philip T Quinlan and Glyn W Humphreys. Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception & psychophysics*, 41(5):455–472, 1987.

[28] Maria Riveiro and Göran Falkman. Supporting the analytical reasoning process in maritime anomaly detection: Evaluation and experimental design. In *2010 14th International Conference Information Visualisation*, pages 170–178. IEEE, 2010.

[29] John Ross, M Concetta Morrone, Michael E Goldberg, and David C Burr. Changes in visual perception at the time of saccades. *Trends in neurosciences*, 24(2):113–121, 2001.

[30] William B Rouse. Models of human problem solving: Detection, diagnosis, and compensation for system failures. In *Analysis, Design and Evaluation of Man–Machine Systems*, pages 167–184. Elsevier, 1983.

[31] John W Senders and Neville P Moray. Human error: Cause. *Prediction, and Reduction, LawrenceErlbaum: Hillsdale, NH*, 1991.

[32] Peter Shepherdson and Jeff Miller. Redundancy gain in semantic categorisation. *Acta psychologica*, 148:96–106, 2014.

[33] David L Sills, Charles Parker Wolf, and Vivien B Shelanski. Accident at three mile island: the human dimensions. 1982.

[34] Daniel J Simons. Attentional capture and inattentional blindness. *Trends in cognitive sciences*, 4(4):147–155, 2000.

[35] Maureen Stone. *A field guide to digital color*. AK Peters/CRC Press, 2016.

[36] Anne Treisman and Stephen Gormican. Feature analysis in early vision: evidence from search asymmetries. *Psychological review*, 95(1):15, 1988.

[37] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

[38] Dal Vernon, C Reising, and Penelope M Sanderson. Ecological interface design for pasteurizer ii: A process description of semantic mapping. *Human factors*, 44(2):222–247, 2002.

[39] Kim J Vicente and Jens Rasmussen. Ecological interface design: Theoretical foundations. *IEEE Transactions on systems, man, and cybernetics*, 22(4):589–606, 1992.

[40] Johan Wagemans, Jacob Feldman, Sergei Gepshtein, Ruth Kimchi, James R Pomerantz, Peter A Van der Helm, and Cees Van Leeuwen. A century of gestalt psychology in visual perception: Ii. conceptual and theoretical foundations. *Psychological bulletin*, 138(6):1218, 2012.

[41] Colin Ware. *Visual thinking: For design*. Elsevier, 2010.

[42] Joel S Warm, Raja Parasuraman, and Gerald Matthews. Vigilance requires hard mental work and is stressful. *Human factors*, 50(3):433–441, 2008.

[43] Jeremy M Wolfe. Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, 1(2):202–238, 1994.

[44] Jeremy M Wolfe and Todd S Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature reviews neuroscience*, 5(6):495, 2004.

[45] David D Woods. The cognitive engineering of problem representations, 1991.

A Appendix: Informed consent form

Informed Consent Form – Master thesis Sita Vriend

The following will provide you with information about this evaluation study that will help you in deciding whether or not you wish to participate. If you agree to participate, please be aware that you are free to withdraw at any point throughout the duration of the evaluation without any penalty.

In this study we will ask you to identify system states using a graphical displays. Your task is to be as accurate and fast as possible to identify the graphical display with the asked system state. There are no known benefits or negative consequences to participating in this evaluation study.

All information you provide will remain confidential and will not be associated with your name. If for any reason during this study you do not feel comfortable, you may leave and receive the agreed upon compensation, no questions will be asked and your information will be discarded. Your participation in this study will require approximately 50 minutes. When this study is complete you will be free to ask any questions. If you have any further questions concerning this study please feel free to contact us through phone or email:

Sita Vriend (master student): Sita.Vriend.0253@student.uu.se or (07 00367433).
Mikael Laaksoharju (supervisor): Mikael.laaksoharju@it.uu.se or (07 39713503)

Please indicate with your signature on the space below that you understand your rights and agree to participate in the experiment.

Your participation is solicited, yet strictly voluntary. All information will be kept confidential and your name will not be associated with any data or findings.

____

City and date

____

Signature

____

Print Name

B Appendix: Information email

Information email

Thank you for showing interest in my study! This email gives you some more information on the study you have signed up for.

You have been signed up to take part on (insert correct date). The study will be held at Pollacksbacken (ITC) in hus 1, room 1313. The study begins at 14.15, please be on time.

I´ve attached a copy of the informed consent form which is yours to keep. You don´t need to print it out, I will provide you a copy to sign when you arrive at the lab.

During the study you will be asked to read system states on a large number of chart on a computer screen. You should perform the tasks as fast and accurate as possible. This information might seem a little abstract now but will become clearer once you perform the experiment. Before signing up, you´ve performed a small pre-test to check if you´re able to complete the study. You will do this test again during the experiment. If you haven´t done the pre-test yet you can do it here: https://pre-test-color.herokuapp.com/

Don´t hesitate to contact me if you have any further questions.

Kind regards,
Sita Vriend

C Appendix: Pre-test survey

Pre-test Survey

Participant Number (on screen): ___________________________________

Participant Number – 72 charts (on screen): ___________________________________

What is your gender?

☐ Male

☐ Female

☐ Other

What is your age?

Age: _______________

How many years did you have English as a subject in school?

☐ 5 of fewer

☐ 6 to 10

☐ 11 or more

Are you currently studying?

☐ Yes, studying program name: _________________________

☐ Yes, studying individual courses at department(s):______________________________________

☐ No, not currently studying

Thanks you! Please click the button on screen to begin the test.

D Appendix: Post-test survey

Post-test Survey

How confident are you that you identified the unique chart on the last four screens?

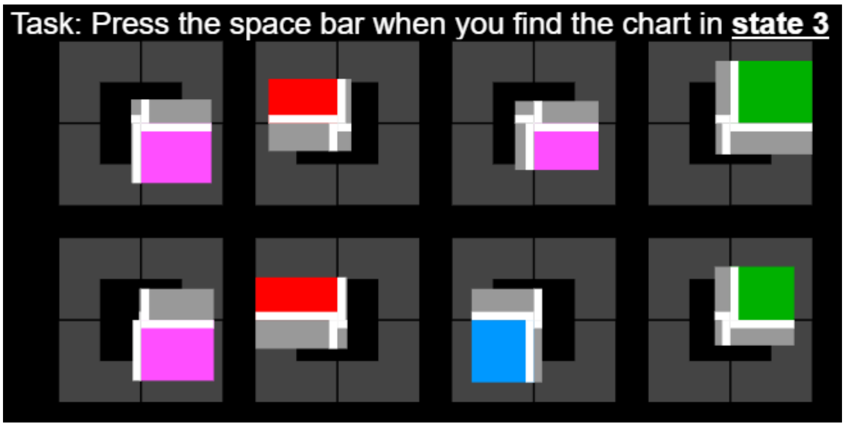
Not at all confident	Not very confident	Neutral	Confident	Very confident
○	○	○	○	○

What was your way of finding the unique chart on the last four screens?
Feel free to write or illustrate in your answer.

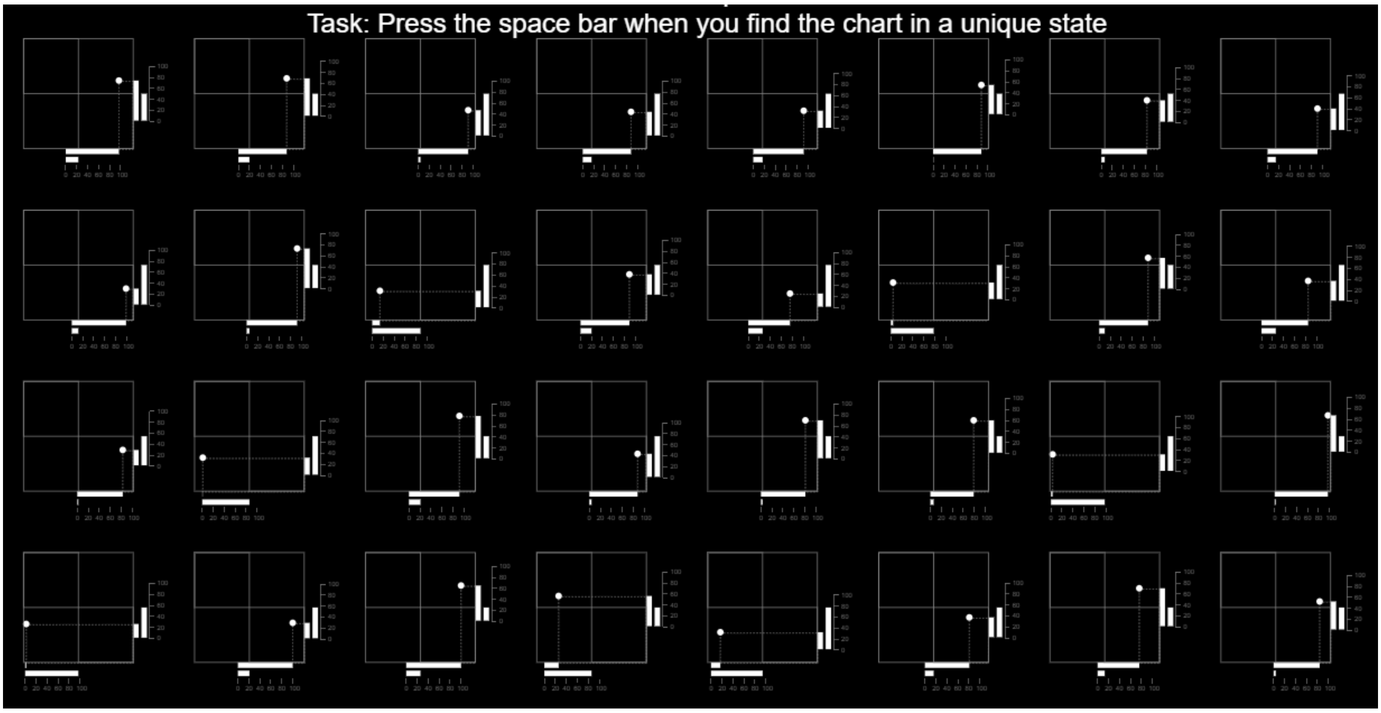
E Appendix: Test session order

Screen	What
Info page	Participant number is given
	Participant is asked to complete the pre-test survey
Color pre-test	Test whether participants can disiguish the colors
Info page	Instructions on how to read the charts including examples
Info page	Instructions on the upcoming task on sequential display training
Training task	Sequential display training
Info page	Instructions on the upcoming task on identifying a display in a known unique state
Training task	Identifying a display in a known unique state
	Participant received feedback on whether the correct display was chosen
Info page	Instructions on the upcoming performance task on identifying a display in a known unique state
Performance task	Identifying a display in a known unique state
Info page	Instructions on the upcoming performance task on identifying a display in an uknown unique state
Performance task	Identifying a display in an uknown unique state
Info page	Link to 72-chart experiment
Info page	Participant number for 72-charts is given
Info page	Instructions on the upcoming task on identifying a display in a known unique state
	Examples are given
Training task	Identifying a display in a known unique state in 72 display screens
	Participant received feedback on whether the correct display was chosen
Performance task	Identifying a display in a known unique state in 72 display screens
Performance task	Identifying a display in an uknown unique state in 72 display screens
Info page	Participant is told the experiment is over
	Participant is told to fill in the post-test survey
	Participant is told to hand in both surveys

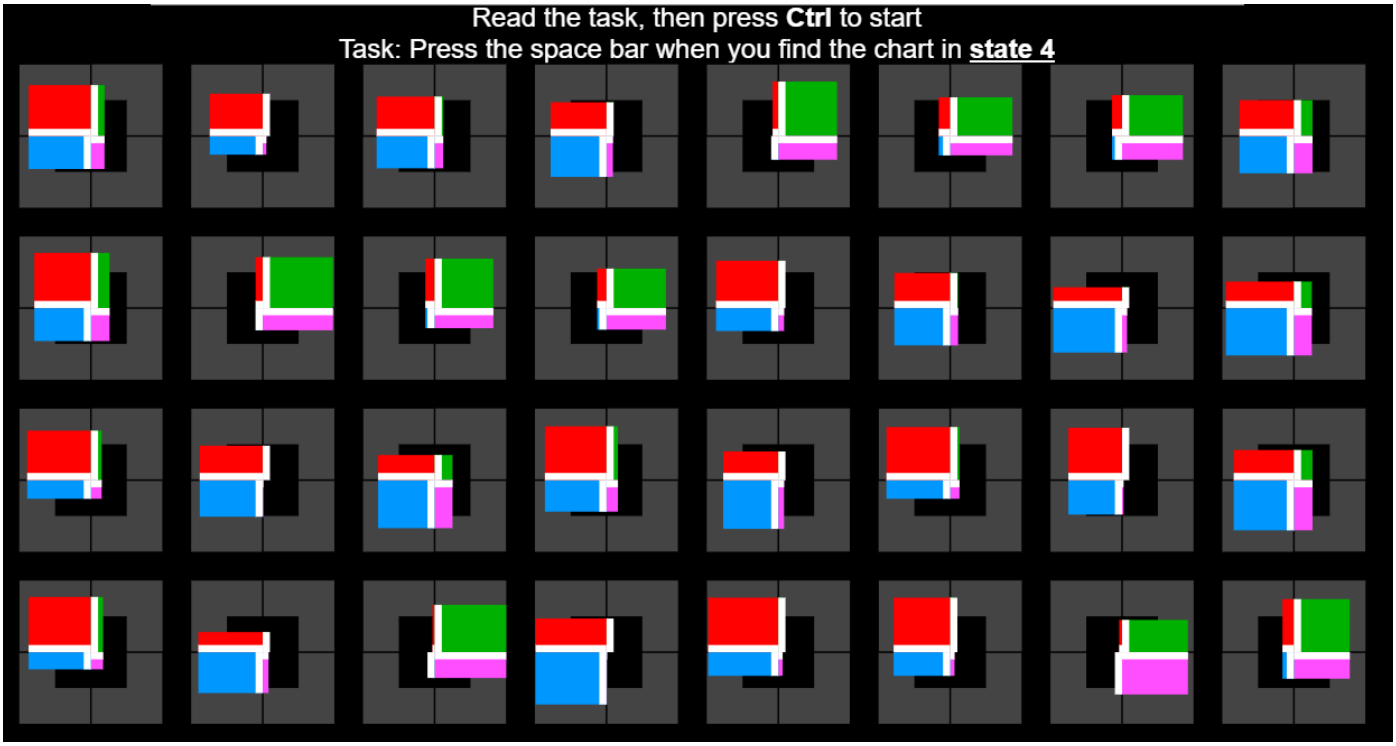
F Appendix: Example screenshot of static 8 displays with a known unique state for the ID1.



G Appendix: Example screenshot of static 32 displays with a unknown unique state for the CCD.



H Appendix: Example screenshot of static 32 displays with a known unique state for the ID4.



I Appendix: Accuracy proportions per person for each combination of dependent variables of static displays

				static display					
				Known state			Unknown state		
				8 displays proportion accurate	32 displays proportion accurate	72 displays proportion accurate	8 displays proportion accurate	32 displays proportion accurate	72 displays proportion accurate
display type	CCD	participant ID	219652	1.00	1.00	1.00	1.00	1.00	1.00
			388524	1.00	1.00	1.00	.75	1.00	1.00
			406553	1.00	1.00	1.00	1.00	1.00	1.00
			409836	1.00	1.00	1.00	.75	1.00	1.00
			419468	1.00	.75	1.00	1.00	1.00	1.00
			491957	1.00	1.00	1.00	1.00	1.00	1.00
			523065	1.00	1.00	1.00	1.00	.50	.00
			559224	1.00	1.00	1.00	1.00	1.00	.75
			691803	1.00	1.00	1.00	1.00	1.00	1.00
			781121	1.00	1.00	1.00	1.00	1.00	1.00
	ID4	participant ID	110740	1.00	.75	.	1.00	1.00	.
			151658	1.00	1.00	.50	.75	.75	1.00
			313833	1.00	1.00	1.00	.75	1.00	1.00
			325959	1.00	1.00	1.00	.50	1.00	1.00
			459272	1.00	1.00	1.00	1.00	1.00	1.00
			491839	.75	1.00	.50	.25	1.00	1.00
			546780	1.00	1.00	1.00	1.00	1.00	.75
			550835	1.00	1.00	1.00	.75	1.00	1.00
			768182	.75	.75	.50	1.00	1.00	1.00
			861964	1.00	1.00	1.00	.75	1.00	.50
	ID1	participant ID	16494	1.00	1.00	1.00	1.00	1.00	1.00
			67462	1.00	1.00	1.00	1.00	1.00	1.00
			73425	1.00	1.00	1.00	1.00	1.00	1.00
			319200	1.00	1.00	1.00	1.00	1.00	1.00
			428752	1.00	1.00	1.00	1.00	1.00	1.00
			493480	1.00	1.00	1.00	1.00	1.00	1.00
			584319	1.00	1.00	1.00	1.00	1.00	1.00
			649136	1.00	1.00	1.00	1.00	1.00	.75
			724868	1.00	1.00	1.00	1.00	1.00	1.00
			880235	1.00	1.00	.75	.00	.00	.00

J Appendix: Accuracy proportion per person for each combination of dependent variables of dynamic displays

				dynamic display					
				Known state			Unknown state		
				8 displays proportion accurate	32 displays proportion accurate	72 displays proportion accurate	8 displays proportion accurate	32 displays proportion accurate	72 displays proportion accurate
display type	CCD	participant ID	219652	1.00	1.00	1.00	.50	1.00	1.00
			388524	1.00	1.00	1.00	.00	1.00	1.00
			406553	.75	1.00	1.00	.50	.50	.00
			409836	.75	1.00	1.00	.00	1.00	.75
			419468	1.00	1.00	1.00	1.00	1.00	1.00
			491957	1.00	1.00	1.00	.50	1.00	1.00
			523065	1.00	1.00	.50	.25	.00	.00
			559224	1.00	1.00	1.00	1.00	1.00	1.00
			691803	1.00	.75	1.00	1.00	1.00	1.00
			781121	1.00	1.00	1.00	.50	1.00	1.00
	ID4	participant ID	110740	1.00	.50	.	.75	.75	.
			151658	1.00	1.00	1.00	.00	.00	1.00
			313833	1.00	1.00	1.00	1.00	1.00	1.00
			325959	.75	1.00	1.00	.25	.75	.50
			459272	1.00	1.00	1.00	1.00	1.00	1.00
			491839	.75	.75	.50	.25	.75	.75
			546780	1.00	1.00	.75	.50	1.00	.75
			550835	1.00	1.00	1.00	.75	1.00	1.00
			768182	1.00	1.00	.75	.75	.25	.75
			861964	1.00	.75	1.00	1.00	1.00	1.00
	ID1	participant ID	16494	1.00	1.00	1.00	.75	1.00	1.00
			67462	1.00	1.00	1.00	.50	.75	1.00
			73425	1.00	1.00	1.00	.75	1.00	1.00
			319200	1.00	1.00	.75	.25	1.00	1.00
			428752	1.00	1.00	1.00	.00	1.00	1.00
			493480	1.00	.75	1.00	.75	1.00	.75
			584319	1.00	1.00	1.00	.00	1.00	1.00
			649136	1.00	1.00	1.00	.25	1.00	1.00
			724868	1.00	1.00	1.00	.50	1.00	1.00
			880235	1.00	1.00	1.00	.50	.00	.00

K Appendix: Median latency per person for each combination of dependent variables of static displays

				static displays					
				Known state			Unknown state		
				8 displays median latency	32 displays median latency	72 displays median latency	8 displays median latency	32 displays median latency	72 displays median latency
display type	CCD	participant ID	219652	1.97	6.75	10.42	11.29	36.40	15.00
			388524	1.67	6.21	3.38	5.17	10.83	11.24
			406553	4.49	23.42	4.70	15.15	24.19	24.54
			409836	1.80	3.77	9.54	4.83	15.13	15.60
			419468	2.99	5.88	28.30	8.18	49.90	24.95
			491957	4.38	16.22	10.99	26.18	42.90	37.93
			523065	4.10	12.19	10.63	8.18	22.61	.
			559224	.65	.89	6.71	.54	.45	16.03
			691803	2.44	3.87	4.26	9.75	14.60	34.97
			781121	1.58	3.47	15.41	6.09	11.61	29.81
	ID4	participant ID	110740	4.14	9.32	.	12.97	13.36	.
			151658	2.03	4.09	22.61	6.47	15.87	57.59
			313833	1.34	4.30	5.80	8.82	6.48	7.37
			325959	1.65	11.35	17.45	.62	15.33	14.77
			459272	1.80	2.40	2.18	12.70	12.90	14.21
			491839	4.90	8.26	16.20	19.23	15.55	19.10
			546780	3.64	8.76	4.98	8.96	8.21	26.27
			550835	1.73	3.98	11.48	4.90	8.95	12.29
			768182	2.94	4.66	4.87	3.03	23.15	18.93
			861964	1.75	3.82	6.49	4.78	4.66	8.65
	ID1	participant ID	16494	.99	.96	1.09	1.00	1.08	.93
			67462	.53	.54	.35	.44	.36	1.91
			73425	2.22	5.08	1.01	2.28	1.29	.72
			319200	1.58	1.87	.81	2.45	3.06	2.45
			428752	2.34	1.73	1.82	1.31	1.59	1.74
			493480	2.57	8.91	2.81	5.11	10.92	8.17
			584319	2.26	6.10	2.25	3.95	4.09	1.91
			649136	.31	.29	.61	.92	1.22	.97
			724868	.75	.92	.98	1.49	1.60	1.72
			880235	3.14	11.13	4.40	.	.	.

L Appendix: Median latency per person for each combination of dependent variables of dynamic displays

				dynamic displays					
				Known state			Unknown state		
				8 displays median latency	32 displays median latency	72 displays median latency	8 displays median latency	32 displays median latency	72 displays median latency
display type	CCD	participant ID	219652	3.58	12.28	21.86	8.33	8.78	24.51
			388524	3.03	10.08	16.32	.	9.43	9.79
			406553	4.48	11.57	19.64	14.75	19.73	.
			409836	8.41	10.26	27.51	.	11.64	12.60
			419468	3.51	9.57	17.50	45.19	89.85	36.98
			491957	4.77	8.98	31.64	21.80	16.83	31.85
			523065	3.68	6.60	15.17	12.21	.	.
			559224	3.21	17.55	21.18	7.16	5.46	20.88
			691803	2.88	11.18	18.93	2.97	8.36	18.93
			781121	3.80	8.53	29.13	7.48	9.84	24.13
	ID4	participant ID	110740	4.58	5.62	.	7.49	8.18	.
			151658	3.88	11.26	20.81	.	.	11.83
			313833	3.46	5.89	6.85	4.59	7.14	10.22
			325959	4.76	6.83	14.66	15.90	8.70	9.02
			459272	4.53	6.88	8.11	8.03	5.80	10.61
			491839	4.56	13.25	6.58	6.01	11.19	5.33
			546780	5.23	5.63	9.15	4.12	6.20	13.12
			550835	4.41	5.42	7.56	6.53	5.43	6.12
			768182	4.18	8.52	8.44	9.84	8.93	10.38
			861964	3.64	4.07	6.90	6.95	5.24	8.03
	ID1	participant ID	16494	2.00	3.08	4.19	2.25	2.72	3.39
			67462	4.78	4.92	4.81	8.97	5.09	5.16
			73425	3.82	4.03	2.92	2.37	2.17	3.59
			319200	3.33	4.72	4.54	.	3.79	3.97
			428752	3.17	3.62	3.43	.	3.51	3.88
			493480	4.59	7.01	7.44	9.30	10.83	8.84
			584319	3.63	10.64	3.56	.	3.95	3.29
			649136	1.89	2.69	2.30	4.32	2.56	2.84
			724868	2.05	2.70	2.58	3.70	2.28	4.39
			880235	6.35	7.63	30.91	.	.	.

M Appendix: Median latency difference in seconds between the 32 displays and the 8 displays per participant for each combination of dependent variables for the static displays where the unique state is known.

				latency difference
display type	CCD	participant ID	219652	4.78
			388524	4.54
			406553	18.93
			409836	1.97
			419468	2.89
			491957	11.84
			523065	8.09
			559224	.24
			691803	1.43
			781121	1.89
	ID4	participant ID	110740	5.18
			151658	2.06
			313833	2.96
			325959	9.70
			459272	.60
			491839	3.35
			546780	5.11
			550835	2.25
			768182	1.71
			861964	2.06
	ID1	participant ID	16494	-.03
			67462	.01
			73425	2.85
			319200	.30
			428752	-.61
			493480	6.34
			584319	3.85
			649136	-.02
			724868	.18
			880235	7.99