STARS

Electronic Theses and Dissertations, 2004-2019

2014

Facilitating Information Retrieval in Social Media User Interfaces

Anthony Costello
University of Central Florida

Part of the Industrial Engineering Commons

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Showcase of Text, Archives, Research & Scholarship

FACILITATING INFORMATION RETRIEVAL IN SOCIAL MEDIA USER
INTERFACES

by

ANTHONY MARK COSTELLO
B.S. Embry-Riddle Aeronautical University, 1998
M.S. University of Central Florida, 2007

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2014

Major Professor: Waldemar Karwowski

ABSTRACT

As the amount of computer mediated information (e.g., emails, documents, multi-media) we need to process grows, our need to rapidly sort, organize and store electronic information likewise increases. In order to store information effectively, we must find ways to sort through it and organize it in a manner that facilitates efficient retrieval. The instantaneous and emergent nature of communications across networks like Twitter makes them suitable for discussing events (e.g., natural disasters) that are amorphous and prone to rapid changes. It can be difficult for an individual human to filter through and organize the large amounts of information that can pass through these types of social networks when events are unfolding rapidly. A common feature of social networks is the images (e.g., human faces, inanimate objects) that are often used by those who send messages across these networks. Humans have a particularly strong ability to recognize and differentiate between human Faces. This effect may also extend to recalling information associated with each human Face. This study investigated the difference between human Face images, non-human Face images and alphanumeric labels as retrieval cues under different levels of Task Load. Participants were required to recall key pieces of event information as they emerged from a Twitter-style message feed during a simulated natural disaster. A counter-balanced within-subjects design was used for this experiment. Participants were exposed to low, medium and high Task Load while responding to five different types of recall cues: (1) Nickname, (2) Non-Face, (3) Non-Face & Nickname, (4) Face and (5) Face & Nickname. The task required participants to organize information regarding emergencies (e.g., car accidents) from a Twitter-style message feed. The messages reported various events such as fires occurring around a

fictional city. Each message was associated with a different recall cue type, depending on the experimental condition. Following the task, participants were asked to recall the information associated with one of the cues they worked with during the task. Results indicate that under medium and high Task Load, both Non-Face and Face retrieval cues increased recall performance over Nickname alone with Non-Faces resulting in the highest mean recall scores. When comparing medium to high Task Load: Face & Nickname and Non-Face significantly outperformed the Face condition. The performance in Non-Face & Nickname was significantly better than Face & Nickname. No significant difference was found between Non-Faces and Non-Faces & Nickname. Subjective Task Load scores indicate that participants experienced lower mental workload when using Non-Face cues than using Nickname or Face cues. Generally, these results indicate that under medium and high Task Load levels, images outperformed alphanumeric nicknames, Non-Face images outperformed Face images, and combining alphanumeric nicknames with images may have offered a significant performance advantage only when the image is that of a Face. Both theoretical and practical design implications are provided from these findings.

This effort is dedicated to my wife, Jing. Thank you for all your love and support. You

are an extraordinary woman, wife and mother.

ACKNOWLEDGMENTS

I do not know where to begin. There are so many to thank and not enough to words. To Mom and Dad, thank you for all your love and support. To my wife Jing, this simply would not have happened without you and I thank you for that. To my chair, Dr. Waldemar Karwowski, thank you for providing a stable foundation so I could soar. To my committee members, Dr. Ahmad Elshennawy, Dr. Peter Hancock, Dr. Richard Gilson, and Dr. Haydee Cuevas thank you for your commitment to helping me reach my full potential. To my colleagues at SA Technologies: Dr. Mica Endsley, Dr. Cheryl Bolstad, Laura Strater, Dr. Jennifer Riley and Fleet Davis, thank you for the unending support over the years. No man is an island.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

As the amount of computer mediated information (e.g., emails, documents, multi-media) we need to process grows, our need to rapidly sort and store information likewise increases. In order to store information effectively, we must find ways to sort through it and organize it in a manner that facilitates efficient retrieval.

Though the average home user may have little need for truly efficient tools for sorting, organizing and storing electronic information, operators in domains such as crisis response, military intelligence analysis or tactical operations centers, must have access to the most efficient means necessary to sort, organize, store and retrieve various types of information. Though specially designed software can be built to deal with some of these challenges, the ever-changing nature of information types and modes of delivery make it difficult for some software platforms to scale well with the speed of developments. For instance, the proliferation of social media networks such as Twitter has created whole new methods of transmitting and receiving information of various types. Twitter allows people to follow topics, other users and conversations about specific subjects (Visualscope, 2012). A unique feature of Twitter is that it only allows people to write messages of up to 140 characters at time. These brief messages are known as 'Tweets'. A Twitter feed source can be an individual or an organization. Unless a Twitter feed is protected, absolutely anyone can locate the feed and 'follow' or subscribe to it. For example, if someone wants the latest headlines from a news source such as Reuters, he or she can simply search for and 'follow' Reuters to receive Reuters' Tweets virtually instantaneously. Since Reuters is a global news organization, their Tweets are usually the

latest headlines from around the world and can include Internet hyperlinks to other Tweets, websites, media, etc. People can post comments on a given Tweet, though they are still limited to the 140-character format. Users are allowed to follow an unlimited number of Twitter feeds and Twitter feeds can have an unlimited number of followers. In this fashion, Twitter is, in some ways, a sophisticated version of the classic old paper ticker tapes that proliferated in the late 1800s and early 1900s. A critical feature of Twitter is the ability to tag a Tweet with a keyword known as a hashtag. For example, if a user had just witnessed an event such as a launch of the Space Shuttle Atlantis, he or she may write: "Just saw the launch. What a beautiful sight! #AtlantisLaunch". The '#' symbol denotes a hashtag which functions as a keyword. Other Twitter users who are not even following the person who wrote that post can still see the post by searching for the hashtag. All Twitter posts using that hashtag will then be displayed almost instantly as well as all future posts containing that hashtag. This can occur regardless of who posted the message, when it was posted or where the source is geographically located. In recent years, the power of the hashtag has become evident through their use in organized demonstrations and uprisings (Segerberg & Bennet, 2011) since it allows users to instantaneously communicate and coordinate with respect to a specific topic. Regardless of the nature of the post, Twitter allows users to connect with one another by allowing a user to directly send messages to the author of a post, though messages are still limited to 140 characters.

The instantaneous and emergent nature of the communications medium provided by Twitter makes it suitable for other events whose nature can be amorphous and prone to

rapid change. Organizations that are responsible for public safety may use Twitter to communicate important public safety announcements to followers. Though the intent of some of the Twitter feeds may be one-way communication, Twitter may still be used as a two-way medium---a lesson the New York Fire Department learned recently. As Hurricane Sandy touched down in the northeast of the United States in October of 2012, a Social Media Manager for the New York City Fire Department had to spontaneously begin relaying the content of direct messages she had received via Twitter to emergency response dispatchers (Moody, 2012) because the messages contained urgent pleas for help from citizens who had no other method to contact authorities due to loss of and/or overload of phone lines. Since there was no other way to contact authorities, citizens had located the FDNY Twitter feed and sent direct messages to the author, in this case, the Social Media Manager. Even though the FDNY discouraged this practice, citizens disregarded the directive because the overload of the phone lines had reached unprecedented levels. For example, New York City Mayor Michael Bloomberg stated that on a typical day dispatchers received around 1000 calls per half hour, but due to Hurricane Sandy, calls to 911 numbered in excess of 10000 per half hour (Saul, 2012). Though the Social Media Manager was only supposed to be using the Twitter feed to relay official announcements to the public, the system quickly became an alternative platform for the FDNY to transfer critical information to and from citizens in need of assistance.

There are four critically important features of this event where Twitter was utilized as the communications medium: 1) events can occur in parallel in that multiple messages can

arrive almost at the same time as opposed to one-on-one phone calls that are received serially in a conventional emergency dispatch environment, 2) the communications from the victim are in text form and 3) due to network latency, typing speed and/or limited Twitter message length (140 characters), multiple messages can come in over time and be mixed with messages from other victims and 4) citizens who send these messages may be identified by a nickname (e.g., SuperCat48, dghf381z) and an image of almost anything (or no image at all). In addition to managing this information, the dispatcher must also communicate either via voice or through electronic means (i.e., text) to other team members or external entities to gather information and/or coordinate to meet relevant goals.

As social networking mediums like the Twitter platform continue to proliferate in our society, it can offer advantages like it did during Hurricane Sandy as an alternative route to communicating with emergency services. Given today's austere budget environments, a high priority is put on doing more with less. Adding more resources in the form of more dispatchers for unusual scenarios like those encountered with Hurricane Sandy may be a practical move from the perspective of making use of the current system, but it does nothing to increase the efficiency of the individual dispatcher. Nor does it address the current inefficiency in the handling of communications received via emerging technologies such as social media networks. According to a member of the Orange County Fire Department in Central Florida, the organization has begun grappling with how to handle information transmitted via social media (J. Mulhall, personal

communication, November 10, 2012) as part of larger scale emergency response exercises.

Assuming these trends continue, it is worth taking another look at how conventional information management paradigms and software can be enhanced to help dispatchers handle large amounts of Twitter traffic in a more efficient manner. In such an environment where organization and retrieval of information is a time critical process, the information management tool utilized to represent and manage the data can have a tremendous impact on the efficiency and productivity of the operator. The unique features of Twitter may also provide avenues to leverage specific aspects of human cognition to reduce load in operational environments where multiple sources of text-based information must be organized, tracked and retrieved under time pressure.

For instance, Twitter is a medium that allows users to use both images and alphanumeric nicknames to identify themselves. Though some alphanumeric nicknames are somewhat mnemonic in nature (e.g., SuperCat78), the rapid proliferation of the medium and/or the desire for privacy often leads to nicknames that are deliberately designed to exclude words (e.g., A75gHle). Twitter users are not required to use images. But if they choose to do so, they may use any image they choose (e.g., personal portrait, abstract art, an inanimate object). The use of personal portraits (i.e., faces) as parts of the Tweet messages creates the potential to use the unique face recognition abilities of the human brain to potentially increase productivity for certain types of tasks.

Therefore, it is the purpose of the dissertation to test the efficacy of human face images to boost operator productivity during an organize-and-retrieve task in an environment where task load is high.

CHAPTER TWO: LITERATURE REVIEW

Human Information Processing

In order to understand how a user interface could be modified to achieve the research goal, it is necessary to review particular aspects of human behavior and cognition. Human information processing is often analyzed with the help of a multi-stage model (Figure 1) (Wickens & Hollands, 2000). During the perception stage, stimuli are perceived through detection, recognition, identification or categorization. Finite and limited attentional resources restrict the number of stimuli that can be processed in this manner. In the cognition stage, the stimuli are processed through both working and long-term memories. Working memory is characterized, in part, by its vulnerability to interference and limited capacity, whereas long-term memory (LTM) serves as the brain's permanent storage system. During the cognition stage, attention resources are required to make comparisons between perceived stimuli and information stored in long-term memory. If a physical reaction is chosen during the cognition stage, it is carried out in the motor response stage. This physical response also draws from attention resources.

Of particular interest to this discussion is how humans recall and recognize information and the cognitive load associated with each. In the context of the design goals, these processes are affected largely by properties of visual perceptual process, attentional resources, working memory, and LTM.
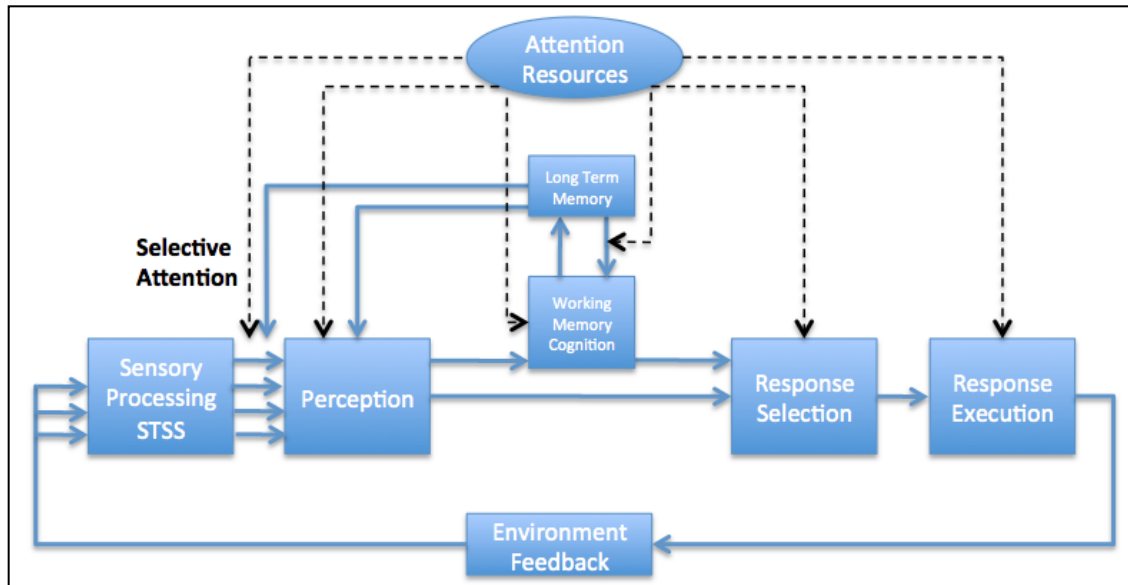
Figure 1 Human Information Processing Model adapted from Wickens and Hollands (2000).

Working Memory

Baddeley and Hitch (1974) developed a model of working memory that is divided into several components (Figure 2): central executive, phonological loop and the visuospatial sketchpad (VSS). Each working memory sub system processes different types of information.

In this model, the *central executive* is an attentional system that includes the capacities to focus attention, divide attention between two important targets/stimulus streams, task switching and long-term memory (LTM) access.

The *phonological loop* handles verbal and acoustic information via the phonological store and the articulatory rehearsal system. The phonological store is a limited-capacity system that is vulnerable to interference. The phonological similarity effect (Baddeley & Hitch,

8

1974) manifests itself in that the *recall* of items (e.g., individual letters) that are similar in sound is poorer than that of dissimilar items, though this effect is absent for similarity of semantic meaning when a standard serial recall paradigm is used (Baddeley, 2002). Though the long-term recall of 10-item sequences shows the opposite effect where recall accuracy seems to be based on semantic rather than acoustic coding. Therefore, as large blocks of phonologically based information need to be remembered, it appears that semantic coding appears to be more effective (Baddeley, 2002). The articulatory loop, when not suppressed, prevents the decay of information in the phonological store via rehearsal (i.e., sub-vocal articulation). Articulatory rehearsal plays an important role in the serial recall of words in that rehearsal helps subjects to remember words before they decay, but the effectiveness of rehearsal drops as word length increases (i.e., word length effect) (Baddeley, 1974). Apparently, a distinction may be made between phonological materials that are presented in an auditory form as opposed to a written form in terms of how they are handled in the phonological loop. Since the phonological similarity effect can be removed by presenting material visually as opposed to aurally, it seems that auditory material goes directly to the phonological store as opposed to written material that must first be processed through the articulatory loop (Baddeley, 2012). Therefore, it appears that attention must be allocated to facilitate sub vocalization of written materials in the articulatory loop.
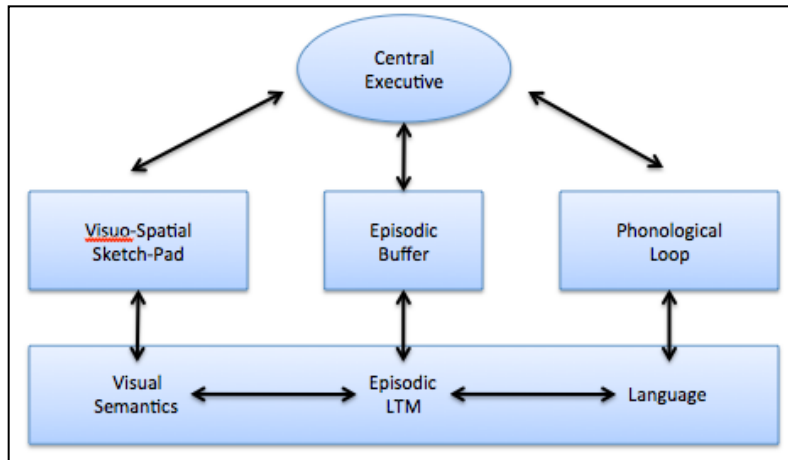
The *visuospatial sketchpad (VSS)* temporarily maintains and manipulates memory for either visual or spatial information (Baddeley, 2002). This sub system helps with spatial orientation and solving visuospatial problems (Logie, 1995). It interfaces visual and spatial information incoming from either LTM or the senses and can bind it with motor, tactile or haptic information as well (Baddeley, 2002). Processes within this sub system are vulnerable to disruption via spatial and visual interference (Baddeley, 2012). Through a series of experiments, Klauer and Zhao (2004) demonstrated that visual and spatial components of the VSS are separable (i.e., dissociated) from each other as well as from the central executive. Therefore, visual short-term memory (STM) and spatial STM appear to be separate functions. In addition, both Visual STM and Spatial STM are limited in capacity (Logie, 1995).

The original model was recently expanded to include a fourth sub system. The *episodic buffer* sub system provides for the integration of information from the other sub systems and the LTM so that it may be actively manipulated and maintained. Baddeley (2002)

describes this sub system as mnemonic in character and serves to bind information from multiple sources that are consciously being considered simultaneously.

This model of working memory provides insight into how different types of information that must be processed are encoded. Ultimately, these codes are manipulated into a form that is meaningful and can then be transferred to LTM (Sanders & McCormick, 1992). Since attention is required to process information, its limits have an impact on the amount and type of information humans can process. The raw information humans perceive can be of various types and formats. Information that is of a type or format that requires more attention to acquire, encode and store may have an adverse impact on overall task performance. Particularly in circumstances where there may be multiple distractions, information that is not properly formatted can draw excessive attentional resources to the point where task failure or undesirable delays may result (Schneider & Shiffrin, 1977). Therefore, it is important to find ways to format information in a manner that results in the least amount of demand from attentional resources.

<u>Memory: Recall and Recognition</u>

Recall simply refers to the ability to retrieve desired information from LTM. Recognition refers to the ability to determine if information presented externally corresponds to information held in LTM. In either case, recall and recognition ability could potentially be enhanced while the information is being encoded via chunking information into appropriate pieces (Miller, 1956; Shiffrin & Nosofsky, 1994), creating mnemonics (i.e., structuring the information in meaningful ways) and/or creating effective retrieval cues (Tulving & Thomson, 1973; Lansdale, Simpson & Stroud, 1990).

11

In terms of organizing information in our daily lives, we often attempt to structure disordered information for the purpose of aiding in retrieval (i.e., recall- and recognition-based searches). For example, if an operator were in charge of analyzing physical versions of documents (e.g., faxes), how would he or she organize them? Initially, the documents would come out of the fax machine and probably be placed on a desk before the operator. As the number of faxes increased, along with the diversity of information they contained, the operator's first instinct would probably be to spread the papers out across a table. This action, referred to by Kirsh (1995) as "spatially decomposing a task" (p. 44), is an initial effort to organize the information without the benefit of context. It is the kind of action that is common to humans in that, in the face of unknown structure and complexity, they attempt to build one (Aaltonen & Leikoinen, 2006). The operator's instinctive step, therefore, can be viewed as an initial effort to use objects in the environment to simplify the task of organization until a more efficient method can be devised. As the operator learns more about the relationships between different documents, piles will inevitably form. Malone (1983) poses four reasons for the creation of such paper piles: "(1) the mechanical difficulty of creating labeled file folders, binders and so forth, especially if multiple levels of classification are desired, (2) the cognitive difficulty of creating appropriate categories and deciding how to classify information in a way that will be easily retrievable, (3) the desire to be reminded of tasks to be done, (4) the desire to have frequently used information easily accessible" (p. 111).

In other words, when confronted with disordered information, humans utilize skills and techniques to decrease cognitive load by first organizing it at a high level—in this

example, spatially. This higher-level organization usually results in a loose, generalized set of clusters (e.g., individual piles of papers), which form a starting point for further analysis (Kirsh, 1995). The next step is to analyze the characteristics of each object, noting both unique and common attributes. An iterative process follows where clusters are reorganized, eliminated or parsed down to smaller clusters. During the organization process and any associated tasks upon which the organization effort depends, humans must constantly recall where objects are or where objects belong. If a human cannot recall where an object is, they must simply rifle through the objects until they recognize the target object. In the context of a time critical operational environment, the notion of spatial decomposition is of use because it describes a method that humans naturally use to reduce cognitive load.

In the context of computer user interfaces, work has been done to increase the effectiveness of recall and retrieval with respect to information organization (e.g., structure) and information representation (e.g., icons, text, icons & text). Lansdale (1988a) examined how users organize their personal information using a typical hierarchical text-based folder paradigm and found that participants recall information better when they have organized and categorized information themselves. In a related study, Lansdale (1988b) had participants organize data using icons tags as opposed to text. Results indicated that icons did not automatically result in high levels of performance. However, he did not view the results as conclusive indication that icons cannot offer superior performance. Rather, the results may indicate that users must be able to create a meaningful association between the information representation (i.e., cues

13

or attributes) and the content within it. In other words, if the cues are meaningful to the participant, it will affect how the information is encoded. He concluded that the appropriate use of meaningful icons in the context of the task should offer superior performance. At the least, visual attributes such as a colored shapes or categorically color coded icons can be used to facilitate partial recall, which serves as a kind of filter. In this manner, the filtering action may serve to reduce cognitive load. In a later study by Lansdale (1990) that compared verbal and pictorial methods to support recall, the results indicated modality was not a significant factor. Rather the results indicated that attributes that support meaningful (i.e., semantic) fit most likely had the greatest impact on recall performance. Altogether, these results did not indicate that pictorial methods for recall were inferior to verbal methods.

Given the nature of how information is organized on the computer screen, it is necessary to examine how the spatial locations of objects serve as a retrieval cue in recall. As discussed earlier, the act of encoding memories exerts demands on already limited attention resources. Additional research indicates that some mental operations may require minimal attention and may be thought of as "automatic" in this sense (Schneider & Shiffren, 1977; Shiffren & Schneider, 1977; Posner & Snyder, 2004). Conversely, other mental operations require more attentional resources and may be considered effortful (Hasher & Zack, 1979). It has been further suggested that such automatic processes may apply to the processing of location information (Jones & Dumais, 1986). It would seem, then, that location information might provide not only an effective retrieval cue, but also be less costly to encode (i.e., exert lower demand on attentional resources)

than other retrieval cues (Hasher & Zack, 1979). Though this a tremendously attractive

prospect for computer-based information retrieval, implementation may prove more

elusive since several factors may influence the automaticity of a process in general such

as consistency of use, necessity for intention, patterns of development, potential for

training-related improvement and susceptibility to attentional decrements (1979).

This perspective is supported by latter research examining whether spatial memory offers

advantages over other modes in the context of computer-based information retrieval.

Jones and Dumais (1986) did a series of experiments examining location-based, name-

based and location-and-name based retrieval. Results indicated that name-based retrieval

proved superior to location-based retrieval and that the combination of name and location

provided only modest improvement. However, Lansdale (1991) pointed out the use of

target object attributes, such as spatial location, depends upon the task at hand and how

the document is used. Furthermore, Lansdale asserts that coding methods such as spatial

location or color are not independent of the overall semantic meaning that is created. In

some situations, a document's location is in a specific location for a specific reason

(much like documents in a physical office environment). In other words, the spatial

location has a meaning that is combined with other important attributes about the

document. These results seem to indicate that specific coding strategies (i.e., use of

retrieval cues) are not necessarily superior to one another outside a specific context of

use.

If this is the case, then the goals and demands of a task are likely important determining factors for the efficacy of specific types of coding strategies. These strategies are often inter-dependent on the types of organizational structures employed for a given task. Effective organizational structures may offer memory mnemonics in their own right. Though some types of mnemonics are verbal in nature they may be visual and spatial in nature as well (see Bellezza, 1981 for a review). Fiore, Johnston and Van Duyne (2004) maintain that mnemonics help users to organize information, assimilate new information with prior knowledge and deepen processing. Niemala and Saariluoma (2003) found that this paradigm may improve the recall of labels and that spatial grouping supports recall of locations for visually organized objects. Though it should be noted that, in their study, blank icons with text labels were used and participants were not allowed to organize the information themselves. Regardless, the authors concluded that *spatial relations* are more likely to be encoded if they are semantically linked in a manner consistent with task demands. The authors also theorized that the spatial groups facilitated chunking (Miller, 1956; Shiffrin & Nosofsky, 1994), but that this form of mnemonic may be more effective when icons are visually grouped based on semantic relationships. Therefore, the use of icon clusters with meaningful spatial arrangements may serve as visual mnemonics that provide patterns that may aid recognition-based search. Varying the presentation of the information may affect how well humans can remember a given item or group of items by providing a mnemonic structure to the data.

Recognition can also be affected by the visual format of the object prior to processing. In other words, humans process verbal and visual codes differently. In the operational

16

context (i.e., Twitter Feeds) of the current study, messages consist of up to 140 characters

formatted into sentence strings. If user must scan the text message for a specific piece of

information, he or she is most likely employing the phonological store and articulatory

loop (i.e., sub vocalization) to process the information (i.e., read) and find a match. In

terms of recognition efficiency, Ellis and Dewar (1979) concluded that the meaning of

existing symbolic (i.e., icon-based) traffic signs are understood more quickly than their

verbal counterparts. Camacho, Steiner and Berson (1990) found that icons produced

faster search and selection times over alphanumeric-based information. It is possible that

the ability to recognize objects that are not alphanumeric based could be used to facilitate

faster information management. Along these same lines, it is further possible that

human's unique abilities to differentiate between faces could be similarly helpful in

boosting performance.

<p align="center">Face Recognition</p>

Humans have a particularly strong ability to recognize human faces as opposed to other

objects. Visual recognition of unfamiliar faces is high: 71% on immediate testing as

compared to 46% for inkblots and 33% for snowflakes, even after a 48 hour delay

(Bruce, 2009). This phenomenon is often referred to as the *Face Superiority Effect.*

Though found to be superior to object recognition performance in certain contexts

(Goldstein & Chance, as cited in Jensen, 1988; Wiseman & Neisser, 1974), face

recognition performance can be sensitive to such simple changes as inverting the face, a

change that can negatively affect performance (Yin, 1969). Performance can also be

negatively affected if only isolated parts of faces are presented or if face parts are

scrambled (Patterson & Baddeley, 1977; Tanaka & Farah, 1993). In the case of a

<p align="center">17</p>

photograph, features of the image such as illumination, direction of illumination and/or

mode of presentation (i.e., video versus still image) can also negatively affect recognition

(Bruce, 2009). Photographs presented as line drawings likewise diminish face recognition

performance (Davies, Ellis & Shepherd, 1978). Bruce (2009) maintains that the ability to

discriminate faces at a fine level is more or less consistent with Biederman's (1987)

model of visual object recognition (recognition-by-components or RBC theory). In this

model, Biederman suggests that all objects are visually recognized using primitive

geometric features referred to as geons, which are used for object recognition at a base

level. Beyond the base level, metric variations and features of surface coloration must be

utilized to recognize objects such as faces (Bruce, 2009).

Any item may be recognized using any combination of the following codes: specific

pictorial details, abstract visual or structural code or verbal code (i.e., names) (Bruce,

2009). These codes can provide further specificity to how an object is recognized. In the

case of faces, different forms of coding can provide deeper levels of detail for

discrimination. Such codes can be of the visually derived semantic type (e.g., sex, race,

age), the expression type (e.g., happy, sad), the identity-specific semantic type such as the

place where they were first seen (i.e., context) and the name type (e.g., John Smith)

(Bruce & Young, 1986; Young et al., 1986).

Face recognition appears to be a mostly visual process (Ellis, 1975), in that verbal coding

does not seem to play a role in the encoding and recall of foundational codes that are

created when humans see a face for the first time. Bruce and Young (1986) described the

following model for the face recognition process: 1) visually perceived structural codes describing the surface features of the face are compared with those stored as face recognition units, 2) a match is the result of a face recognition unit being triggered, and 3) a match allows access to identity-specific semantic and name codes. Following this model, during the face encoding process, the visual structural codes are formed first, followed by the formation of expression and visually derived semantic codes. After that, identity-specific semantic and name codes may be formed. Identity-specific and name codes may require some level of familiarity to encode (e.g., a famous face) or a context within which to remember the face (e.g., the place where a person was first seen) (Bruce & Young, 1986). In the context of retrieval cues that may be provided to facilitate face recognition, Bruce and Young (1986) suggest that encoding specificity theory (Tulving & Thomson, 1973) may partially explain how context cues facilitate retrieval if they provide the same context that was used during encoding.

Prior research with deaf participants (Arnold & Murray, 1988) has indicated that the deaf were able to discriminate faces more accurately than non-deaf participants and non-deaf participants who knew sign language. Those non-deaf participants who knew sign language did outperform other non-deaf participants, however. It is possible that these results occurred for any of the following reasons: deaf participants had naturally further developed their visuospatial sketchpad; deaf participants did not rely in any way on their phonological loop (i.e., they did not repeat common adjectives used to describe faces such as hair color or complexion as they searched for a match); deaf participants had developed some alternative strategy to discriminate between faces; non-deaf participants

19

that knew sign language had developed their visuospatial sketchpad to some extent as they learned and used sign language or some combination of these reasons. Therefore, it is possible that, over-reliance on alphanumeric information that must be processed through the phonological loop has resulted in underutilization of the visuospatial sketchpad for some recognition processes that might otherwise be less costly in terms of attentional resources. For these reasons, it may be possible to leverage human's exceptional face recognition abilities to facilitate search and retrieval in specific operational contexts where speed and accuracy are paramount.

<u>Common Information Organization Paradigms</u>

In the operational context of an emergency operations center, tactical operations center or a crisis response center, humans essentially organize information in a manner that strikes a balance between what can be remembered and what can be recognized rather quickly. If it were possible to remember everything all the time, then the human would not need tools. Our tools are designed to make us more efficient and effective whether it be chopping wood, flying a plane, or helping us sort through huge amounts of information and retrieve what we need as quickly as possible (Norman, 1988). For this particular application, we are focused on how well a person can organize and retrieve information under time pressure.

The efficacy of organization, storage and retrieval processes may be influenced by many factors such as the operational task(s), the type of organizational strategy used to store the objects, how well the organizational strategy is matched with the objects to be organized, how the objects are managed over time and how well the organizational strategy

compliments retrieval behaviors. In the context of email management, organization strategies generally fall into two categories: preparatory and opportunistic (Whittaker, 2011). Preparatory strategies, though time consuming, may help users retrieve emails more efficiently because it utilizes a pre-conceived structure for organizing incoming information (2011) as opposed to opportunistic management strategies. Opportunistic strategies are generally loose or non-existent, which may save time when emails are first received because users do not have to spend time considering where emails should be stored. However, this strategy can cost time later when the user must spend time searching for a target email. Though powerful, preparatory strategies are not always superior. For instance, if changes in retrieval requirements cause a mismatch between the organizational strategy and documents organized within it, time can be lost sorting through a structure that is no longer efficient for retrieval (2011).

In a time pressured environment where the content of messages can take unexpected forms, as is often the case with Twitter, an extensive preparatory organization strategy may not prove to be an efficient or practical means to manage information. This type of problem has existed for a long time in the information retrieval space, starting with menu-driven systems where items are organized by category. In those systems some objects fit into overlapping and fuzzy categories that were hard to classify (Dumais & Landauer, 1983) and sometimes led to inefficient menu structures.

<div align="center">Common Information Retrieval Paradigms</div>

Once a user has been prompted by something in his or her environment to retrieve an object (e.g., document, email), the user must utilize whatever organizational structures

<div align="center">21</div>

they have created and/or the tools at their disposal to find the target object. If the information is completely unorganized, the user must simply rifle through the available objects until the appropriate object is located. This process is typically a recognition-based scanning process (Lansdale, 1988a) where attributes of the desired information are serially and systematically compared with the available objects until the target is found. However, if the information has been organized according to some systematic and relevant structure, then the user may first consider where the desired target object may be in a structure. This strategy uses a kind of recall-based scanning process (1988a). Depending on the type of structure and the amount of information that is contained within it, the user may be able to recall where the target object could be with relative precision. Otherwise, the user may at least recall an appropriate starting point in the organizational structure that is close but not precise. At that point, the user may then begin the recognition-based scanning process to locate and match the retrieval cue to the target. Therefore, for the purposes of the current study, two types of retrieval are recall-based and recognition-based that work more or less in harmony with one another to locate a target object (Fertig, Freeman & Gelertner, 1996; Lansdale, 1988; Jones & Dumais, 1986).

In most cases, a user in a time-pressured environment would be forced to use both recall-based scanning and recognition-based scanning. In most commonly used software applications such as email browsers and Internet browsers, recognition-based scanning processes can be aided by the use of hierarchical folder paradigms in that users do not necessarily need to remember specifics since they can traverse a folder system visually

22

until appropriate target folders are recognized. Though the hierarchical folder structure must use well-conceived organizational structures to be useful and efficient. However, even if this were the case, the serial-nature of recognition-based scanning processes may still increase the time necessary to locate the target document if the available options are extensive. This is particularly important when objects consist primarily of text because there are few overtly distinguishing features at first glance, so text must be read in order to differentiate it from the target object (Czerwinksi, van Dantzich, Robertson & Hoffman, 1999). Given that recognition-based scanning processes may be time consuming, the user may attempt to use recall-based scanning processes as much as possible prior to resorting to the recognition-based scanning. Therefore, users may attempt to use carefully conceived organizational structures (Whittaker, 2011) and mnemonics (Tulving & Thomson, 1971) to bridge the gap and make the overall retrieval task more efficient. Unfortunately, carefully conceived organizational structures that are commonly used in software applications are built using hierarchical folder paradigms that have the disadvantage of essentially being nested lists. This structure does leverage some recall-based scanning processes, but also eventually leads to text-heavy lists of information that have limited sorting features such as: sender, subject, date received (Jones et al., 2005). This leads the user right back to recognition-based processes that may be further slowed by having to serially sort through pure text documents in search of the target information (Lansdale et al., 1989).

Therefore, it may be possible to increase overall user performance in terms of response time and accuracy (i.e., number of errors when asked to retrieve an item) if a user interface is designed in such a manner that it optimizes retrieval processes.

To meet these design goals for this operational context, the following research question should be addressed: How can recall-based and recognition-based scanning processes be accelerated without sacrificing accuracy?

<div align="center">Proposed Prototype</div>

The prototype graphical user interface (GUI) for this study leverages the common information organization and retrieval paradigms to enhance an individual's ability to organize, store, and retrieve electronic information. This GUI incorporates a design feature that allows a more expansive organizational field so that all relevant visual cues (e.g., faces, icons, text) can be viewed at the same time on-screen. This feature also adds a spatial dimension to how objects are organized and stored on the interface. As discussed earlier, this type of design feature may have an impact on how well the information is encoded (c.f. Robertson et al., 1998). It is also possible that spatial attributes are robust in that they may not be negatively affected by some types of phonological load during the encoding phase (c.f. Banbury, Jones & Emory, 1999; Jones, Farrand, Stuart & Morris, 1995; Jones, Madden, & Miles, 1992). Since the GUI characteristics are not being manipulated in the present experiment, the feature should not create noise. Furthermore, using traditional folder hierarchy paradigms like those that are common in email browsers will serve to obscure many of the face images, so a more open format like that of the prototype GUI is desirable. Finally, the open format may

speed organization of the items since the participant will not be forced to create, manage

and manipulate hierarchical folder structures. The prototype GUI will be described in

further detail in the next section.

CHAPTER THREE: METHODOLOGY

Experiment Design

The present study used a within subjects design to determine if the use of human face images as visual representations of Tweet messages would result in greater accuracy and speed over other representation types during a retrieval task. Retrieval performance was examined under continuously increasing levels of Task Load.

Independent Variable (IV)-1 [Within]: Information Representation Type

IV-2 [Within]: Task load

Dependent Variable (DV)-1 [Performance]: Retrieval response accuracy

DV-2 [Performance]: Retrieval speed

DV-3 [Cognitive Load]: Subjective Task Load

Table 1 Experiment design depicting the relationship of each level of each IV

		Task Load		
		Low	Medium	High
Information Representation Type	Nickname			
	Non-Face			
	Non-Face & Nickname			
	Face			
	Face & Nickname			

Information Representation Type (IV-1): This within-subjects variable refers to the type of information representation format that was used for each trial. The representation format refers to the specific visual information that is attached to each Tweet message. This IV has five levels: Nickname, Non-Face, Face, Non-Face & Nickname and Face &

26

Nickname. For example, in a given trial, each Tweet Message was associated with a face. But during another trial, each Tweet Message was associated with a non-face image.

Task Load (IV-2): This within-subjects variable refers to the imposed Task Load conditions under which participants completed each trial. Load was manipulated by varying the number of events and the number of corresponding Tweet messages that arrived in the participant's timeline for each trial. The IV levels were set at Low, Medium and High. For all levels, Tweet messages arrived every 12 seconds and each trial lasted exactly 5.2 minutes. For example, since a total of three Tweet messages were required to cover the three elements of an event (i.e., event type, event address and number injured), this meant that Low Task Load trials had a total of three events that collectively included 9 total Tweet messages. The Tweet messages started arriving at the beginning of the trial with each Tweet message delivered at a 12- second interval until the last message was delivered. The trial did not end upon delivery of the last message, rather it ended when the 5.2 minute interval was finished. Once the messages stopped arriving, the participant had the balance of the trial time to organize the messages. For the Medium Task Load trials, there were fifteen messages comprising five events. For the High Task Load trials, there were twenty-one messages comprising seven events. In this manner, the Task Load level was increased through increasing the number of messages managed and decreasing the amount of time remaining to organize all of them.

Post-Trial Query Retrieval Accuracy (DV-1): After each trial, a participant was presented with one query about one of the events that occurred in the trial, with three parts to the

query. Each event had three elements to be tracked: event type (e.g., fire), event address (e.g., 123 Main St), and number of injuries reported. To answer the query, the participant had to recall each of the three elements and fill them into the blank space provided. They were not timed, but they were not allowed to move on to the next trial without filling in a response to each of the three elements. They were told that guessing was permitted.

Post-Trial Response Speed (DV-2): Though the post-trial queries did not have an imposed time limit, the participant's time to respond to each query was measured in milliseconds. The participants were not told that time was a factor in their responses.

Perceived Mental Workload Measure (DV-3): the Simplified Subjective Task Load Assessment Technique (S-SWAT) questionnaire (Luximon & Goonetilleke, 2001) was administered after each trial to assess time load and mental effort load. In addition, the mental effort question was asked after the recall performance query (DV1) to determine the level of mental effort exerted to answer that question as opposed to executing the task.

<div align="center">Hypotheses</div>

H1: Use of face & nickname combination retrieval cues will result in higher performance in terms of post-trial retrieval accuracy for recall queries than non-face & nickname, non-face, face or nickname Information Representation Types ($\mu_{nickname} < \mu_{non\text{-}face\ image} < \mu_{non\text{-}face\ image\ \&\ nicknames} < \mu_{face} < \mu_{face\ \&\ nickname}$).

Rationale (H1): It is possible that the ability for humans to encode and recognize the unique features of an individual human face, along with specific contextual information such as events associated with a human face, that the face may be used as a kind of visual mnemonic which may increase retrieval accuracy. Since previous research has indicated that name codes may also be encoded with faces, it is possible that name codes will help participants differentiate between faces.

H2: Use of the Face & Nickname retrieval cues will result in higher performance in terms of time to respond to post-trial queries than Non-Face & Nickname, Non-Face, Face or Nickname retrieval cues ($\mu_{nickname} < \mu_{non\text{-}face\ image} < \mu_{non\text{-}face\ image\ \&\ nicknames} < \mu_{face} < \mu_{face\ \&\ nickname}$).

Rationale (H2): It is possible that since humans have a high degree of discrimination accuracy between faces and the faces may trigger contextual information associated with a face, that this ability may also increase performance in terms of speed during the post-trial query response task as opposed to other types of information representations.

H3: Use of the Face & Nickname retrieval cues will result in less perceived mental workload than Non-Face & Nickname, Non-Face, Face or Nickname information representation types ($\mu_{nickname} < \mu_{non\text{-}face\ image} < \mu_{non\text{-}face\ image\ \&\ nicknames} < \mu_{face} < \mu_{face\ \&\ nickname}$).

Rationale (H3): Since human face recognition abilities have been shown to be superior to some other types of visual object recognition, it is possible that participant's subjective opinions of their mental workload will indicate that the use of human Face & Nickname combinations decreases load as compared to other information presentation types.

Participants

Thirty-four participants (males = 18, females = 16) at or over the age of 18 were recruited from the University of Central Florida (UCF) student population. They received a small monetary incentive for their participation. Participants were allowed to quit at any time. They were asked not to take screenshots or notes during the trials so as not to gain an unfair advantage when answering the post-trial questions. Treatment of these participants was in accordance with the ethical standards of the American Psychological Association (APA). No prior experience in crisis management was necessary.

Materials

A brief demographic questionnaire (Appendix B) containing questions such as age, gender, first spoken language, social media use and relevant employment (e.g., employment in an emergency operations center) as administered to each participant prior to the experiment. None of the participants had indicated that they had ever worked in an Emergency Operations Center.

Each Tweet message (Figure 3) contained an image of some kind, depending on the condition. In the *Nickname* condition, an illustration of an egg was displayed as this is the egg illustration currently used for Twitter accounts whose owners choose not to use a particular image. For the *Non-Face* and *Non-Face & Nickname* condition, a color image

of a piece of fruit was used. For the *Face* and the *Face & Nickname* conditions, a portrait-style head-and-shoulders-level color image of a face was used. All images had proportional dimensions of 50x64 pixels at a resolution of 72dpi. An off-white background of each image category (e.g., faces) was the same across all images. Images were not repeated across trials.

The Tweet messages (Figure 3) were accompanied nicknames that consisted of between six and ten character alphanumeric text strings rendered in Arial Font. Nicknames utilized upper- and lowercase letters A thru Z, and integers 0 thru 9.
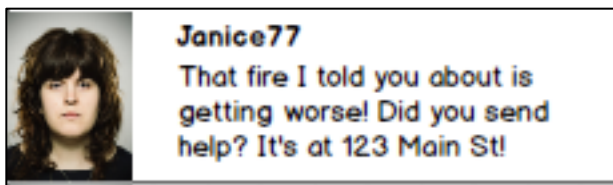
Figure 3 Sample Tweet Message

A total of 15 trials were developed to match each level combination between the five levels of IV1 (Information Representation Type) and three levels of IV2 (Task Load). Each Tweet message consisted of a short (i.e., 140 characters or less) text string written in common American English without abbreviations or acronyms sometimes used on social media. Though shorthand is commonly used in social media (e.g., BTW is widely considered to be equivalent to 'By The Way'), its use in the present experiment may have potentially confounded results and created undesirable noise. However, commonly known location-related acronyms such as: AVE (avenue) or BLVD (boulevard) were used. Examples of Tweet messages that are consistent with this operational context include: 'I see smoke coming out of a home across the street!' or 'There's a nasty three

car pile-up over here. Send help!!' Each Tweet message addressed an event, such as a

fire. But it would take three messages (sent from the same individual) to gather all the

necessary details for an event. As a group, the three messages would convey (1) what the

event was, (2) the event address, and (3) the number of injured people. For example, the

first Tweet message for an event might come from a person with the nickname SuprHead

and simply state he sees a fire. Later, another message would arrive from SuprHead

stating that the fire was at 123 Main St. Some time after that, another message from

SuprHead would state that he see's two people near the fire banged up and coughing.

Messages across multiple events were mixed on the timeline so as not to have the three

messages arrive consecutively for a given event. Each event Tweet message group was

randomly assigned, *a priori*, into one of four types of emergency event: gas leak, car

accident, fire or medical.

During the trial, the participant was instructed to read Tweet messages as they arrived in

the timeline (Figure 4, Arrow 1). When the participant saw a message that was reporting

an event, they would click on the message and press the 'Create Event' button (Figure 4,

Arrow 2). This action put the identity of the individual that is reporting the information

onto the field in the form that matches the trial condition (Figure 4, Arrow 3). For

example, if the participant were in the *Face* condition, then the face image of the

individual who wrote the Tweet message would be in the field (Figure 4, Arrow 3). Or, if

the participant were in the *Face & Nickname* condition, then the face image and the

associated nickname would be in the field (Figure 4, Arrow 4), etc. If the user clicked on

an item in the field, then the entire Tweet message would appear in the Event List (Figure

4, Arrow 5). As more Tweet messages arrived in the timeline, the participant had to determine if any of the messages are associated with events that are being tracked. If the participant determined that a Tweet message in the timeline belonged to an event that was being tracked, then the participant could use the 'Add to Event' button to join the messages together. The newly joined message would appear below the 'parent' message in the Event List (Figure 4, Arrow 6). Each time the participant selected an item in the field, both the original 'parent' message and any associated messages for that event would appear in the 'Event List' (Figure 4, Arrow 5). If the participant changed his or her mind, the message joining and event creation processes could be reversed.

As the Tweet messages arrived during a given trial, the participant was instructed to collect messages related to an individual event and fill out an 'Incident Report ' form (Figure 4, Arrow 7). The form had a field for each of the three event elements (i.e., event type, event address and number injured. The field inputs had dropdown lists pre-loaded with the relevant information for events in that particular trial. The participant was told that filling in the form was a critical part of the job so that a remote dispatcher could determine how emergency response units should respond to each event. The intent of this task was boost engagement with the event information in a meaningful way and to help prevent memory transience (Shacter, 1999). The trial was over when the time 5.8-minute time limit was reached. The time limit was devised to give a participant sixty additional seconds to organize the information after the final message arrived at the High Task Load levels.

After each trial, a post-trial questionnaire (Appendix C) designed to measure perceived mental workload (S-SWAT; Luximon & Goonetilleke, 2001) was administered. The metric included two questions. One question asks participants to rate the mental difficulty of the task on twenty-point scale from "very easy" to "very difficult". The second question asked how much time pressure they felt during the task on a twenty-point scale from "low" to "high".
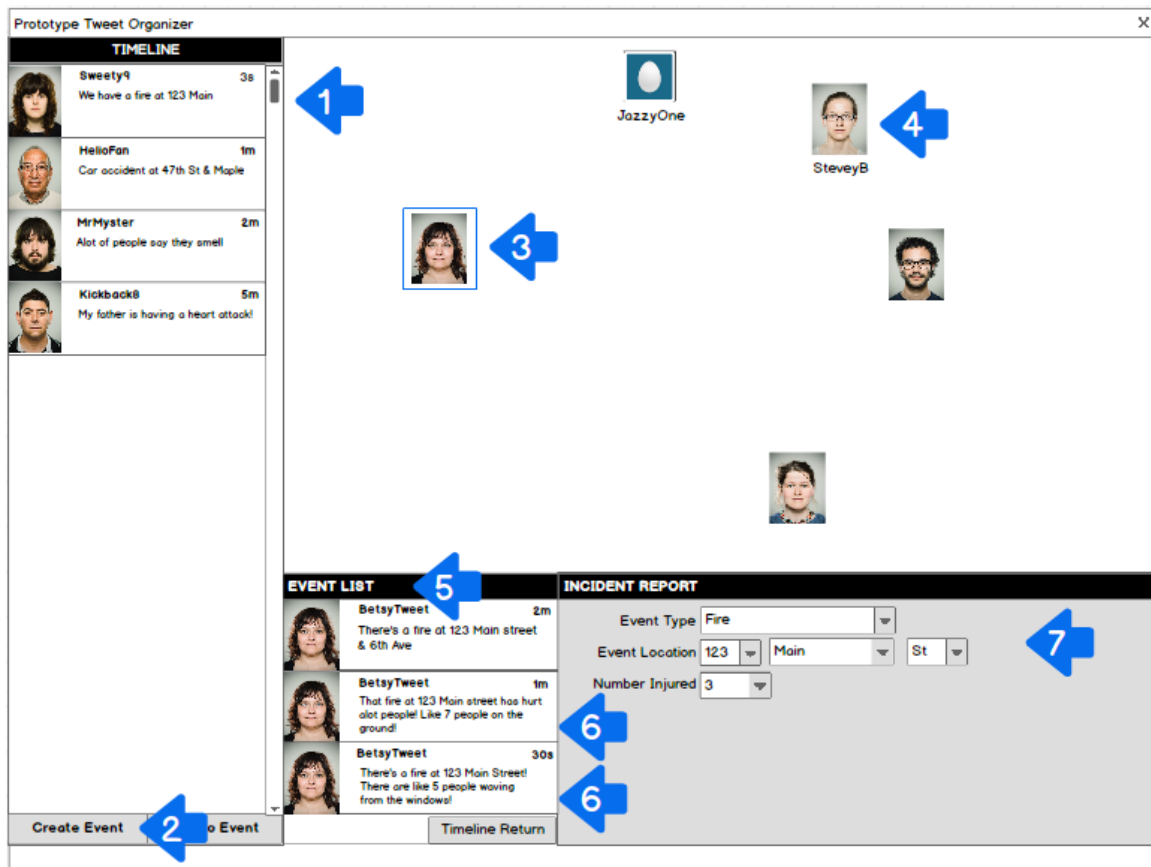
Figure 4 Prototype Tweet Organizer screenshot.

After the post-trial subjective Task Load questionnaires were administered, a post-trial recall query was administered without time limit to determine how well the participant could recall the three information elements associated with each event. The participant was shown the identity of a sender (again depicted in accordance with the condition), but

this time the participant was asked fill in the blanks for the associated event (i.e., event type, event location and number of injuries).

<center>Apparatus</center>

The software prototype was made available to the participant via its own dedicated website. The software required a conventional Internet browser, a standard keyboard and a mouse. The platform was written by the author in C# and ran on the Silverlight framework. It sent data to a secure back-end SQL server.

<center>Procedure</center>

Participants were recruited via UCF's online recruitment system (via SONA at the UCF Institute for Simulation and Training). This study received appropriate IRB approval (Appendix D). The participant was instructed to log into the website to access the experiment. The site briefed the participant about the experiment and asked them to read the informed consent form (Appendix A). After they read the consent form, and provided they felt comfortable with the experiment, they were prompted to confirm their acceptance before they were allowed to proceed. They were then asked to fill out a short demographic questionnaire (Appendix B).

Prior to the first trial, participants received a briefing about the experiment and read a short illustrated tutorial describing expectations for their performance and how to use the software prototype. The participant was given the option to view the tutorial as many times as desired before beginning the experiment. All 15 trials were randomized automatically by the system and checked to ensure that the resulting randomized trial sequence had not been used before. If it had been used, the system would re-roll the

<center>35</center>

randomizer algorithm until a unique sequence was generated. The participant would then proceed through the trials in accordance with the randomized sequence. After each trial, the participant was asked to fill out the aforementioned post-trial subjective Task Load questionnaire (Appendix C), followed by the post-trial trial recall query and its companion subjective Task Load query.

After completion of the experiment, the participant's data was saved both locally in a text file and on a back-end SQL server database platform accessed automatically by the prototype via the Internet. As a redundancy, the participant was asked to save and email a data text file at the end of the experiment. The file and email also served as proof that the specific individual completed the experiment and was used to award credit/compensation. However, the data itself did not have any specific link to that individual.

CHAPTER FOUR: RESULTS

Recall scores and S-SWAT (subjective workload) scores were analyzed. An alpha level of .05 was used for all statistical analyses. Prior to each analysis, the data were screened for normality and outliers.

<u>Recall Scores</u>

One recall question was administered for each of the 15 trials. The intent of the question was to determine if a participant could recall specific information pertaining to one of the events they managed from the trial they completed. The question required three answers: event type, event location (i.e., street address) and number of injuries associated with that event. The participant was not allowed to move on to the next task without answering all three parts of the question. Participants were told that guessing was permitted. If the participant answered all three parts correct, they received a score of 1. If they did not get any of the answers correct, they received a score of 0. If they got any one of the three parts of the question correct, they received a third of a point for each part.

After it was determined that the Recall score data was normally distributed and Mauchly's test for sphericity was not violated, a 3x5 repeated measures ANOVA was performed. The means and standard deviations for recall scores are provided in Table 2.

Table 2 Descriptive statistics - mean recall scores and associated standard deviations for each Information Representation Type across each level of Task Load

Task Load	Nickname	Non-Face	Non-Face & Nickname	Face	Face & Nickname
Low	.549(.294)	.441(.303)	.47(.247)	.568(.239)	.431(.301)
Medium	.284(.234)	.323(.265)	.313(.216)	.225(.268)	.421(.287)
High	.166(.188)	.352(.271)	.352(.258)	.205(.201)	.254(.201)

There was significant main effect of Task Load, F (2, 66) = 27.04, p <.001. The means

and standard deviations for each level of Task Load are depicted in Table 3 and Figure 5.

Task Load	Est. Mean Recall Scores
Low	.492(.033)
Medium	.314(.028)
High	.267(.022)
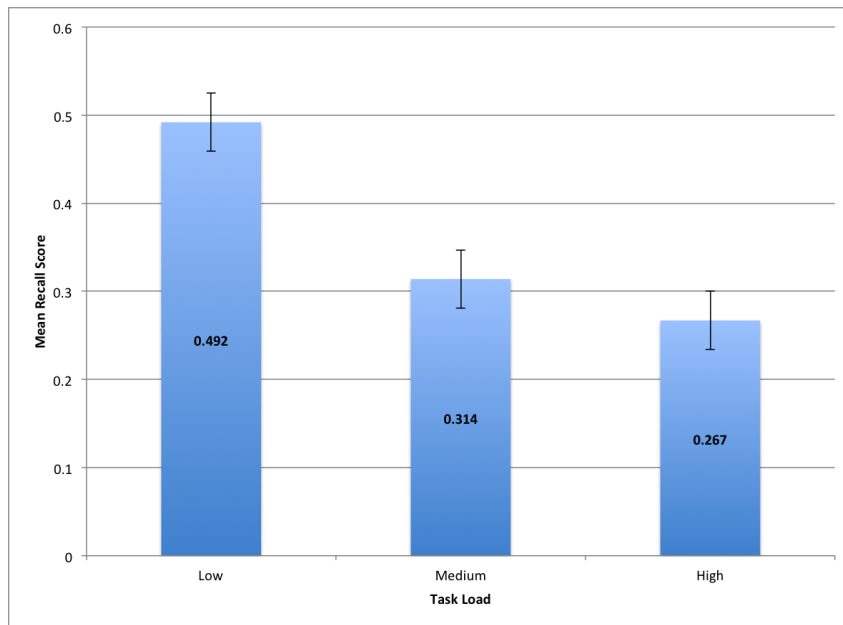
Figure 5 Estimated mean recall scores for each level of Task Load.

Pairwise comparisons (calculated using the Bonferonni adjustment) for each level of

Task Load are depicted in Table 4. There was significant difference in mean Recall score

deltas between low and medium Task Load levels (M = .178, SE = .034, p < .001).

Likewise, there was a significant difference in mean Recall score deltas between low and

high Task Load levels (M = .225, SE = .035, p < .001). The difference in mean Recall

score deltas between medium and high Task Load levels, however, was not significant.

Task Load Levels	Mean Difference	Significance
Low to Medium	.178(.034)	$p < .001$
Low to High	.225 (.035)	$p < .001$
Medium to High	.047(.027)	$p < .275$

The main effect for Information Representation Type was not significant $F(4, 132) = 1.025$, $p < .397$. However, there was a significant interaction (Table 5, Figure 6, Figure 7) between Task Load and Information Representation Type, $F(8, 264) = 5.32$, $p < .001$.

Table 5 Summary of all Mean Recall scores and associated standard error for each level of Task Load and each level of Information Representation Type.

Task Load	Nickname	Non-Face	Non-Face & Nickname	Face	Face & Nickname
Low	.549(.051)	.441(.052)	.471(.042)	.569(.041)	.431(.052)
Medium	.284(.04)	.324(.046)	.314(.037)	.225(.046)	.422(.049)
High	.167(.032)	.353(.047)	.353(.044)	.206(.035)	.255(.035)
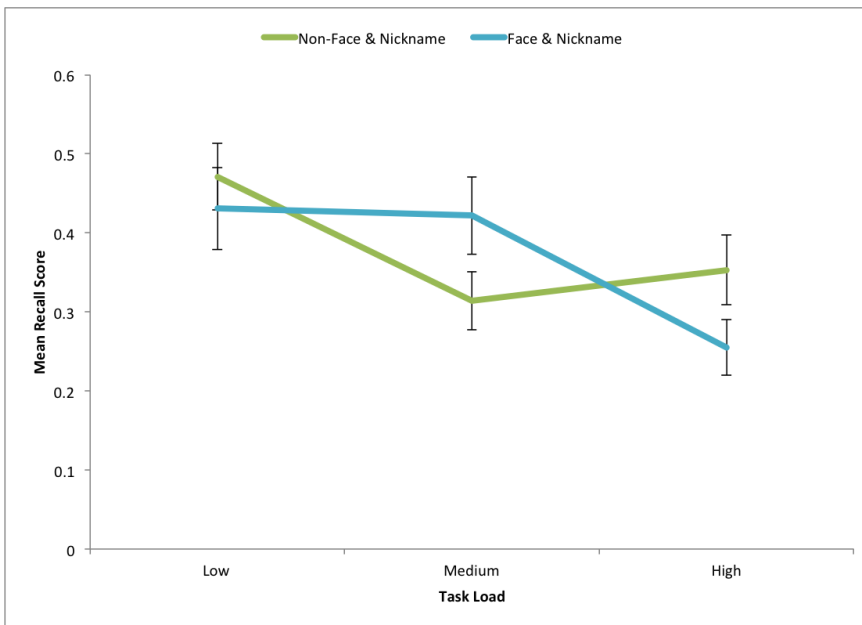
Figure 6 Mean Recall scores plotted against each level of Task Load for Non-Face & Nickname versus Face & Nickname Information Representation Types
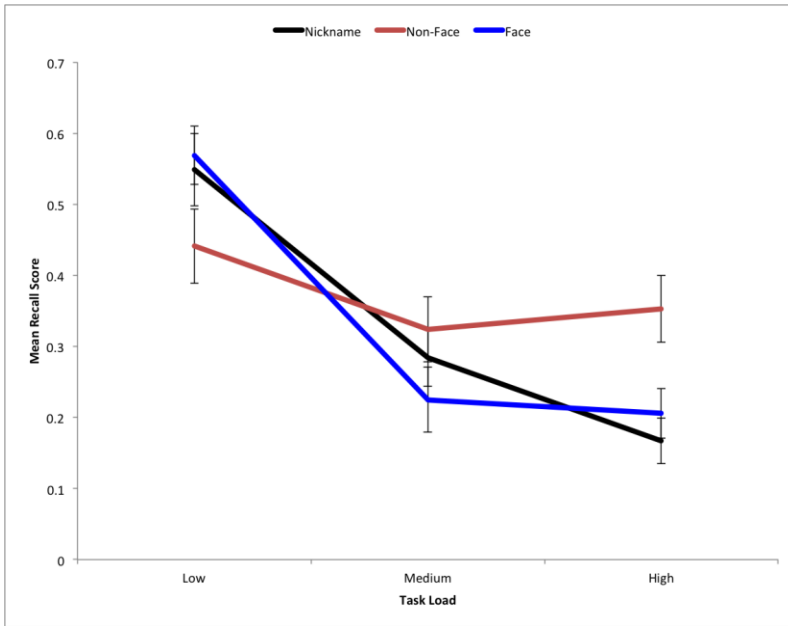
Figure 7 Mean Recall scores plotted against each level of Task Load for Nickname, Non-Face and Face Information Representation Types

Pairwise comparisons between Information Representation Types across each level of Task Load, (i.e., simple effects) are depicted in Table 6. It appears that in the Low Task Load condition participants scored significantly higher in the Face condition vs. Non-Face and Face & Nickname. Though they also scored significantly higher in the Nickname condition vs. Non-Face. However, in the Medium Task Load condition, participants scored significantly higher in the Face & Nickname condition vs. Nickname, Non-Face & Nickname and Face. In the High Task Load condition, participants scored significantly higher in the Non-Face condition over Nickname and Face. Furthermore, participants scored significantly higher in the Non-Face & Nickname condition over Nickname and Face.

Table 6 Significant pairwise comparisons of Mean Recall scores for each level of Task Load and each level of Information Representation Types

Task Load	Information Representation Type	Mean	SD	Mean Diff	Std Error	Sig
Low	**Nickname**	**.549**	**.295**	**.108**	**.046**	**.025**
Low	Non-Face	.441	.304			
Low	**Face**	**.569**	**.240**	**.127**	**.049**	**.013**
Low	Non-Face	.441	.304			
Low	**Face**	**.569**	**.240**	**.137**	**.058**	**.024**
Low	Face & Nickname	.431	.302			
Medium	**Face & Nickname**	**.422**	**.288**	**.137**	**.053**	**.014**
Medium	Nickname	.284	.234			
Medium	**Face & Nickname**	**.422**	**.288**	**.108**	**.046**	**.025**
Medium	Non-Face & Nickname	.314	.216			
Medium	**Face & Nickname**	**.422**	**.288**	**.196**	**.056**	**.001**
Medium	Face	.225	.269			
High	**Non-Face**	**.353**	**.271**	**.186**	**.049**	**.001**
High	Nickname	.167	.188			
High	**Non-Face**	**.353**	**.271**	**.147**	**.055**	**.011**
High	Face	.206	.201			
High	**Non-Face & Nickname**	**.353**	**.259**	**.186**	**.051**	**.001**
High	Nickname	.167	.188			
High	**Non-Face & Nickname**	**.353**	**.259**	**.147**	**.063**	**.026**
High	Face	.206	.201			

For the Non-Face information representation types (i.e., Non-Face vs. Non-Face & Nickname) there was very little difference in performance (Figure 8) if a nickname was present as indicated by non-significant contrasts. Therefore, it appears that the addition of a nickname to the Non-Face image had little impact on performance differences between different levels of Task Load.
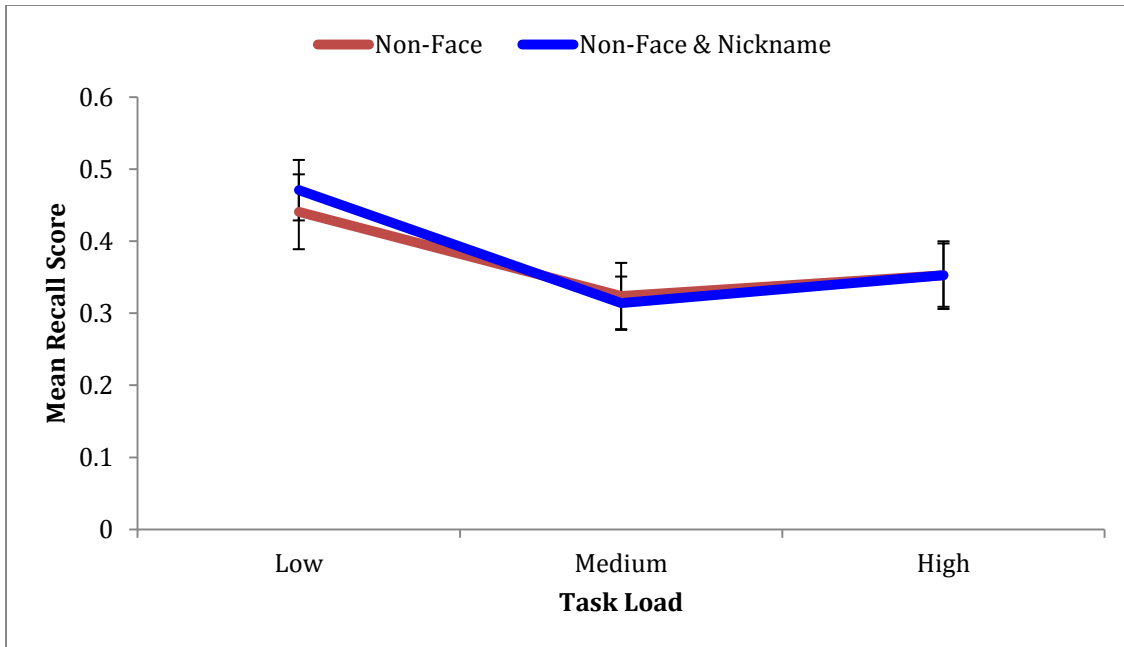
Figure 8 Mean recall scores for Non-Face and Non-Face & Nickname conditions for all levels of Task Load

However, for the Face Information Representation Types (i.e., Face vs. Face & Nickname), contrasts indicate that the performance differences were significant between different levels of Task Load (Figure 9). For Low vs. Medium Task Load, the performance difference for Face vs. Face & Nickname was significant $F(1,33) = 16.5$, $p < .001$, though Face initially underperformed Face & Nickname at Low Task Load, Face & Nickname performed better at Medium and High Task Loads. As Face & Nickname outperformed Face at Medium and High Task Load levels, the performance difference between them was also significant $F(1,33) = 4.169$, $p < .049$. Therefore, it appears that for Medium vs. High levels of Task Load, the addition of a nickname to the Face image resulted in a significant performance boost.
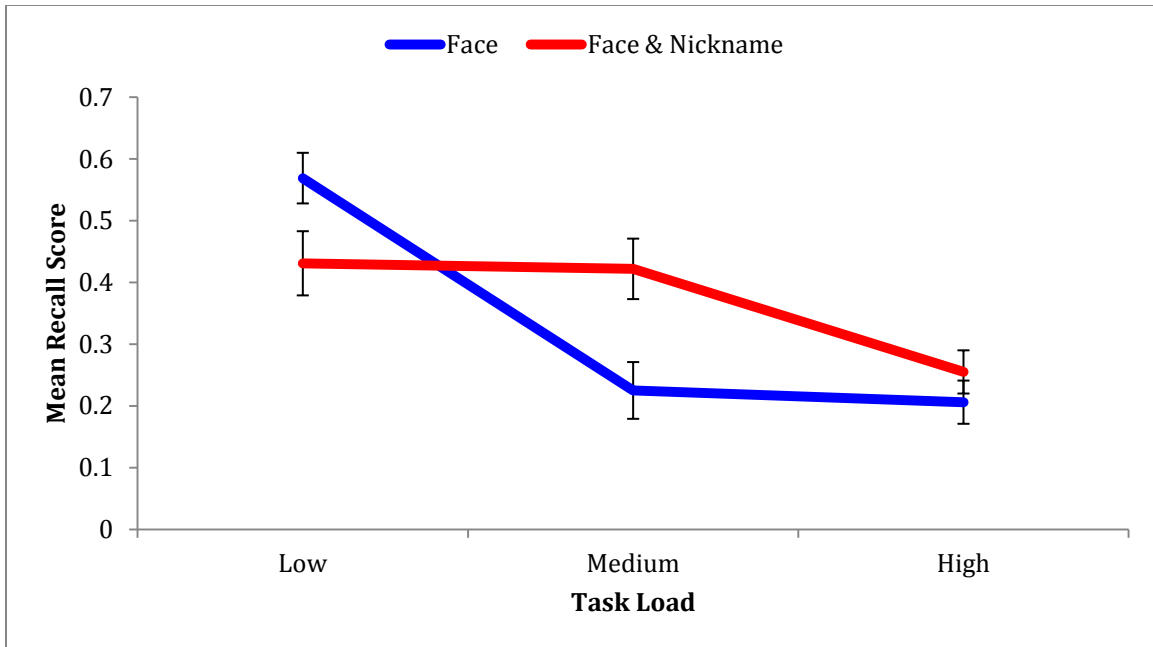
Figure 9 Mean recall scores for Face and Face & Nickname conditions across all levels of Task Load

For the image-only representation types (i.e., Face vs. Non-Face), with respect to the means (Figure 10), Non-Face underperformed Face at Low Task Load, but consistently outperformed Face at Medium and High Task Load. Additionally, contrasts indicate that there was a significant difference for Low vs. Medium Task Load $F(1,33) = 9.981$, p < .003 and Low vs. High Task Load $F(1,33) = 2.562$, p < .001.

Figure 10 Mean recall scores for Face and Non-Face conditions across all levels of Task Load

When comparing Nickname vs. Non-Face & Nickname (Figure 11), not only are the means higher for Non-Face & Nickname at the Medium and High levels of Task Load, but the performance difference between the two from Medium to High Task Load is significant as well $F(1,33) = 5.347$, $p < .07$. Though it should be noted that Non-Face & Nickname did underperform Nickname at Low Task Load. Therefore, it appears that the addition of a Non-Face image to a Nickname resulted in both a substantial performance increase and a performance difference as Task Load shifts from Medium to High.
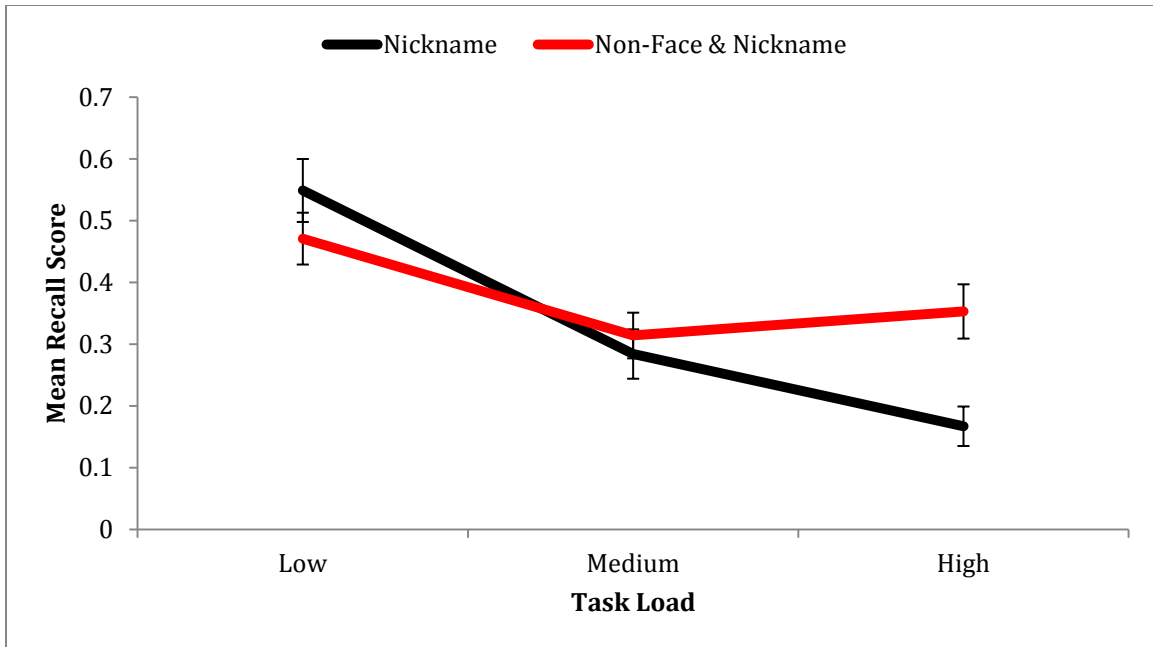
Figure 11 Mean recall scores for Nickname and Non-Face & Nickname conditions across all levels of Task Load

Likewise, when comparing Nickname to Face & Nickname (Figure 12), contrasts indicate that performance differences were significantly different for Low vs. Medium Task Load $F(1,33) = 14.227$, $p < .001$ and Low vs. High Task Load $F(1,33) = 6.685$, $p < .014$. The difference between Medium and High Task Load is not significant, which may indicate that Face & Nickname consistently outperformed Nickname as Task Load shifted from Medium to High.
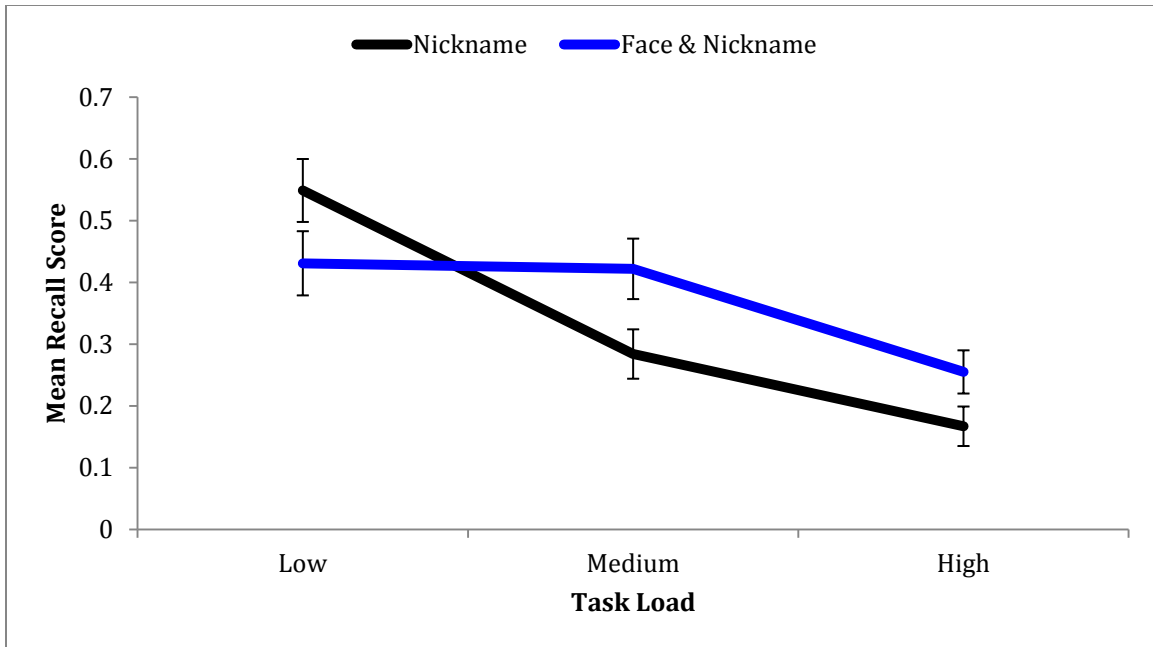
Figure 12 Mean recall scores for Nickname and Face conditions across all levels of Task Load

When comparing Face & Nickname vs. Non-Face & Nickname (Figure 13), contrasts indicate that performance differences were significant at each level of Task Load. There was a significant difference for Low vs. Medium Task Load $F (1,33) = 4.708$, $p < .037$ and Medium vs. High Task Load $F (1,33) = 8.912$, $p < .005$.
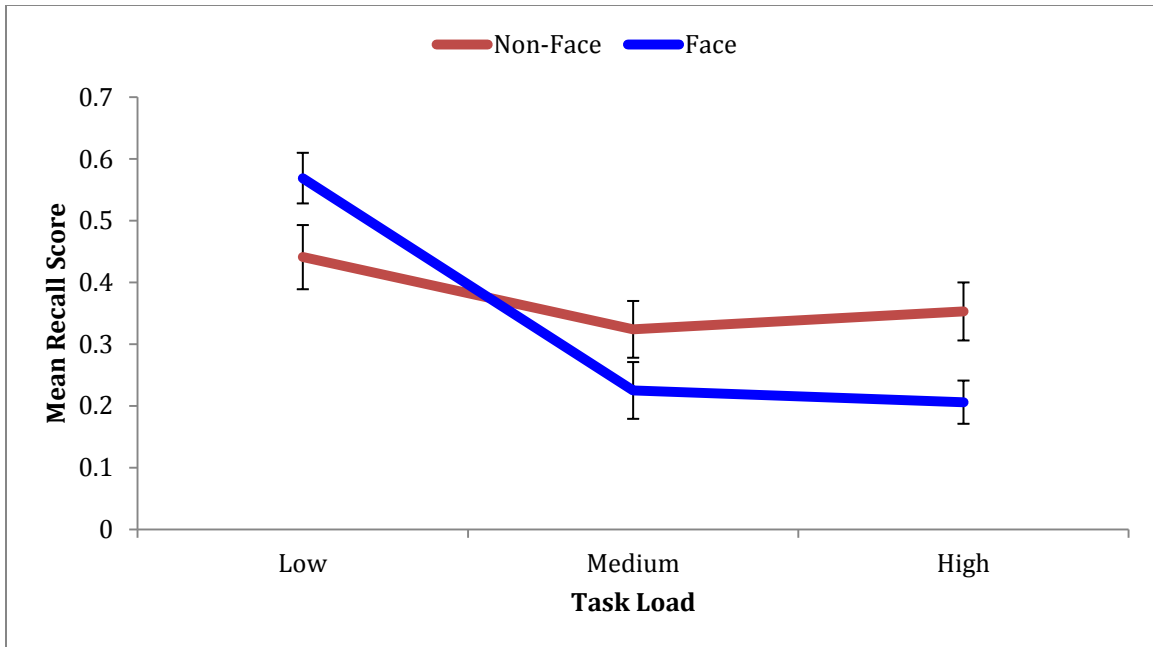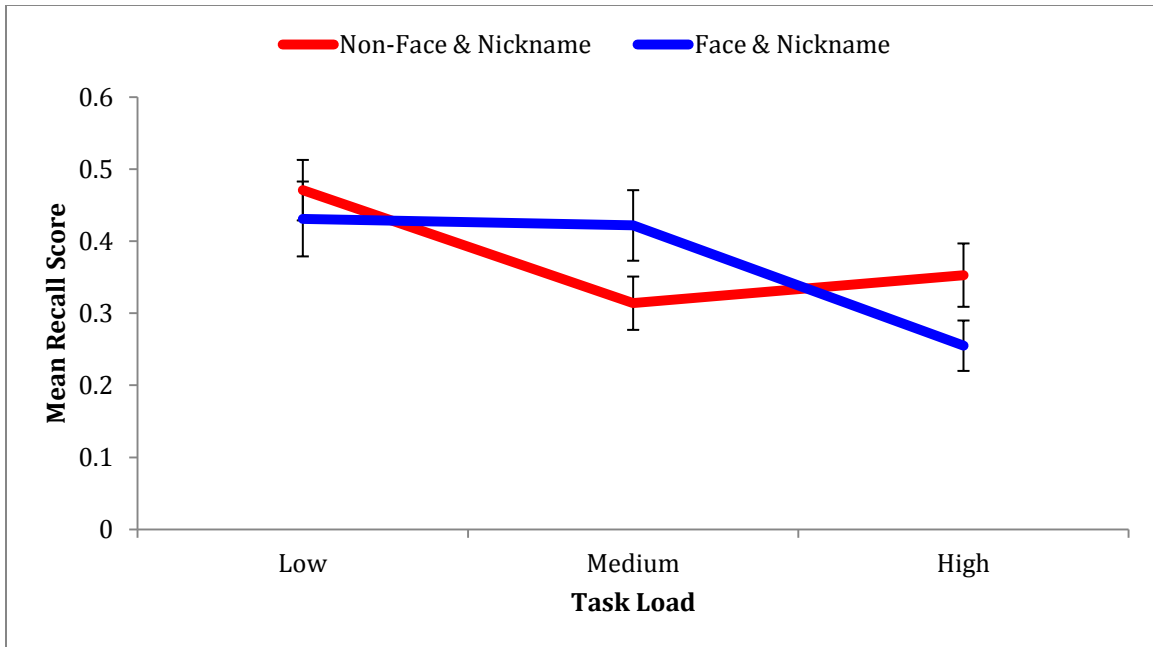
Figure 13 Mean recall scores for Non-Face & Nickname and Face & Nickname conditions across all levels of Task Load

<u>Subjective Task Load Scores</u>

A total of three subjective Task Load questions were asked per trial for each participant.

All questions used the SWAT format (Appendix C). Two of the three questions followed

immediately after the trial. The first one (SWAT-Effort) required the participant to rate

the level of effort they exerted during the task and the second one (SWAT-Time) required

the participant to rate how much time pressure they felt during the task. The third and

final question (SWAT-Recall) immediately followed the recall query and asked the

participant to rate how much effort they exerted to respond to the query. For all of these

metrics, lower scores mean less subjective Task Load.

After it was determined that the SWAT-Effort query response data was normally

distributed and Mauchly's test for sphericity was not violated, a 3x5 repeated measures

ANOVA was performed. There was a significant main effect for Task Load, $F (2,66) =$

25.34, $p < .001$ (Figure 13). Pairwise comparisons (calculated using the Bonferonni

adjustment) of the different Task Load levels (Table 7) indicate that mean SWAT-Effort

response deltas differed significantly between Low and Medium Task Load levels. The

effect continued for the Medium vs. High levels and also for the Low vs. High levels.
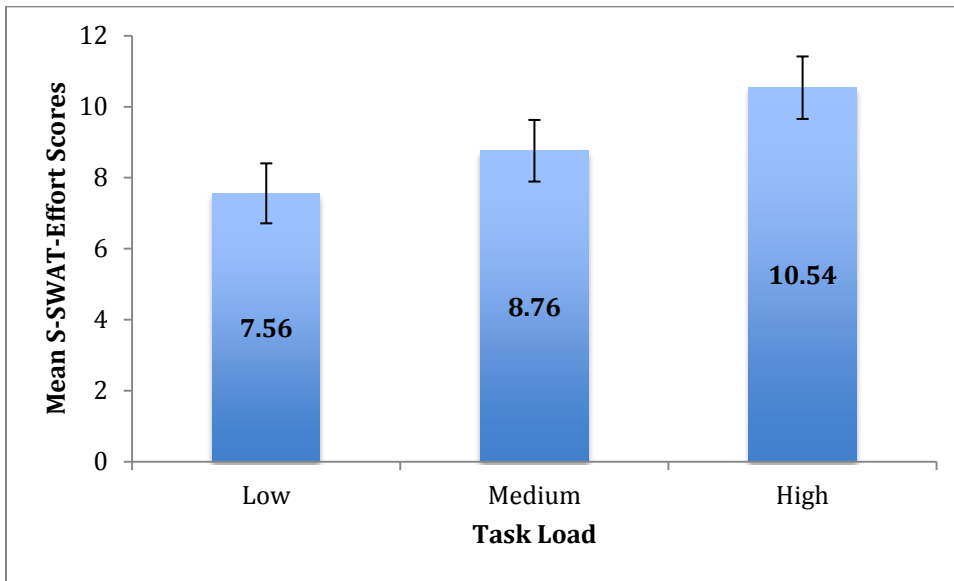
Figure 14 Subjective Task Load (SWAT-Effort) estimated marginal means for each level of Task Load across Information Representation Types

Table 7 Pairwise comparisons for Subjective Task Load (SWAT-Effort) for each level of Task Load across Information Representation Types

(I) Task Load	(J) Task Load	Mean Difference (I-J)	Std. Error	Sig.	95% CI LB	95% CI UB
Medium	Low	1.200	.445	.033	.077	2.323
High	Low	2.976	.514	.001	1.679	4.273
High	Medium	1.776	.261	.001	1.117	2.436

Additionally, there was a significant main effect for Information Representation Type, F

$(4, 132) = 24.608$, $p < .006$) (Figure 14). Pairwise comparisons (calculated using the

Bonferonni adjustment) of the different Information Representation Type levels (Table 8)

indicate that mean SWAT-Effort response deltas differed significantly between

Nickname vs. Non-Face levels. Responses deltas differed significantly between Face vs.

Non-Face levels.
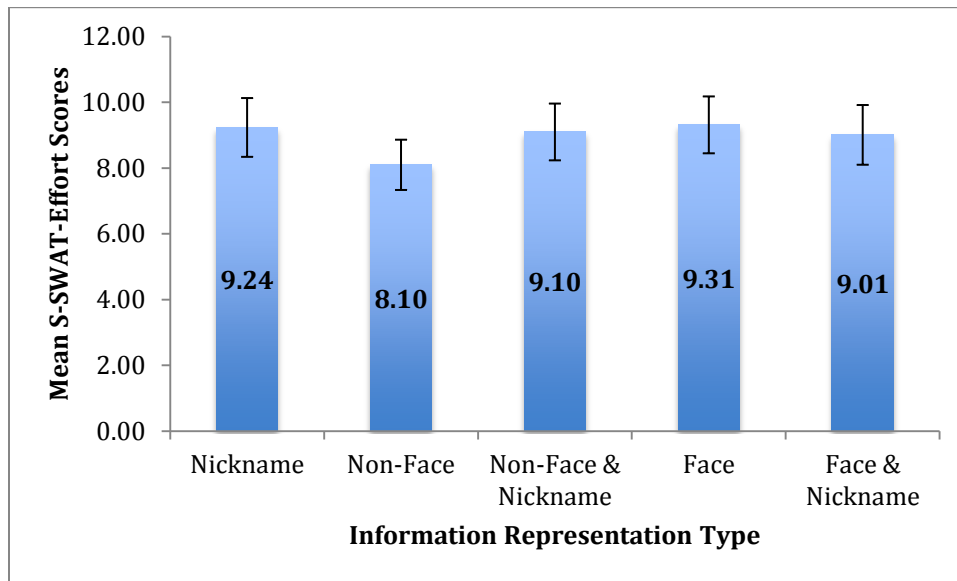
Figure 15 Subjective Task Load (SWAT-Effort) estimated marginal means for each Information Representation Type across Task Load levels

Table 8 Subjective Task Load (SWAT-Effort) pairwise comparisons for each Information Representation Type across Task Load levels

(I) Information Rep. Type	(J) Information Rep. Type	Mean Difference (I-J)	Std. Error	Sig.	95% CI LB	95% CI UB
Nickname	Non-Face	1.137	.37	.042	.023	2.251
Face	Non-Face	1.216	.304	.003	.3	2.131

After it was determined that the SWAT-Time query response data was normally

distributed and Mauchly's test for sphericity was not violated, a 3x5 repeated measures

ANOVA was performed. There was a significant main effect for Task Load F (2, 66) =

25.38, $p < .001$ (Figure 15). Pairwise comparisons (calculated using the Bonferonni

adjustment) of the different Task Load levels (Table 9) indicate that SWAT-Time

response deltas differed significantly between Low vs. Medium Task Load levels. The

effect continued for the Medium vs. High levels and Low vs. High levels.

For the SWAT-Time query, there was a significant main effect for Information Representation Type $F(4,132) = 34.627$, $p < .002$ (Figure 16). Pairwise comparisons (calculated using the Bonferonni adjustment) of the different Information Representation Type levels (Table 10) indicate that mean SWAT-Effort response deltas differed significantly between Non-Face & Nickname vs. Non-Face levels. Responses deltas differed significantly between Face vs. Non-Face levels.
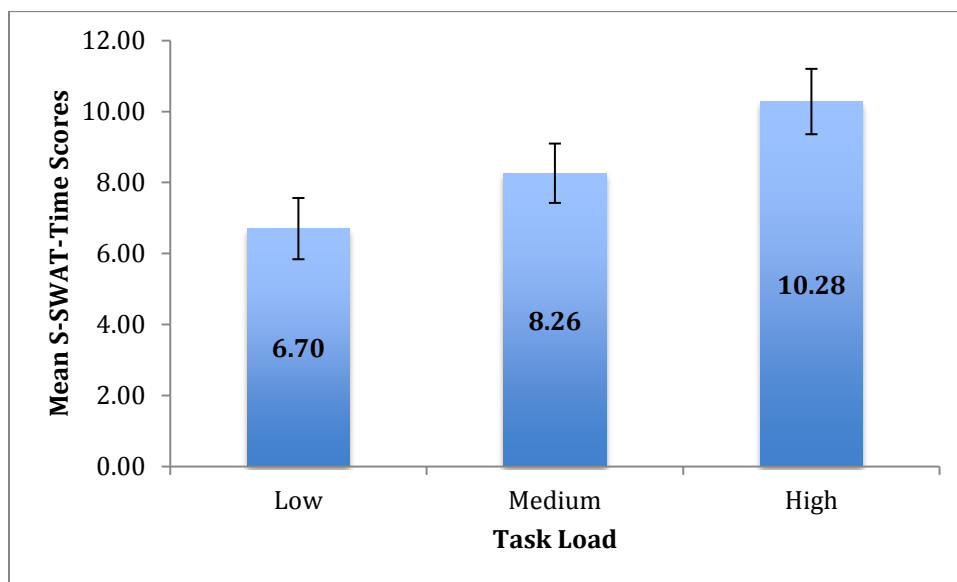
Figure 16 Subjective Task Load (SWAT-Time) estimated means for Task Load across Information Representation Types

Table 9 Subjective Task Load (SWAT-Time) pairwise comparisons for each Task Load level across Information Representation Types

| (I) Task Load	(J) Task Load	Mean Difference (I-J)	Std. Error	Sig.	95% CI LB	95% CI UB
Medium	Low	1.559	.412	.002	.519	2.599
High	Low	3.576	.64	.001	1.963	5.19
High	Medium	2.018	.425	.001	.946	3.089
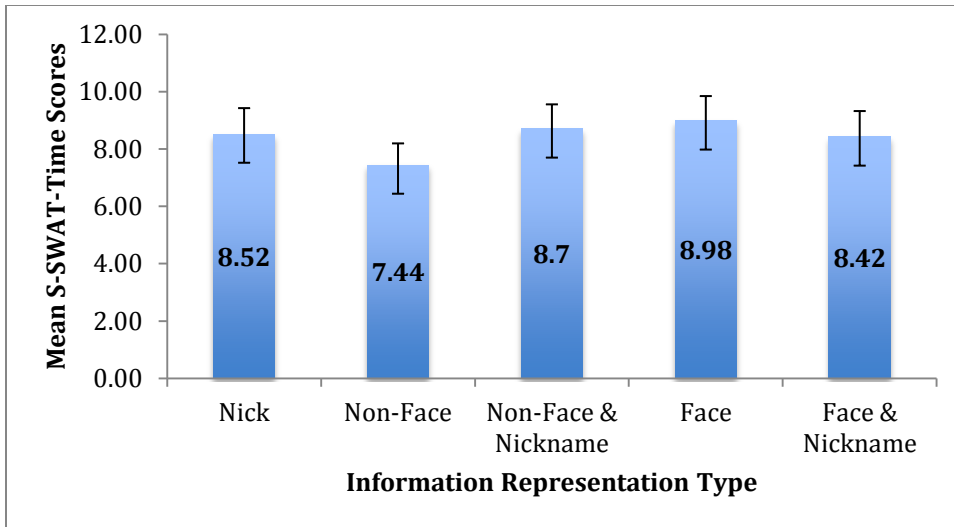
Figure 17 Subjective Task Load (SWAT-Time) estimated means for Information Representation Types across Task Load levels

Table 10 Subjective Task Load (SWAT-Time) pairwise comparisons for Information Representation Types across Task Load levels

(I) Information Rep. Type	(J) Information Rep. Type	Mean Difference (I-J)	Std. Error	Sig.	95% CI LB	95% CI UB
Non-Face & Nickname	Non-Face	1.255	.382	.024	.107	2.403
Face	Non-Face	1.539	.392	.004	.361	2.717

For the SWAT-Recall query (Table 11), a Wilcoxan-Signed Rank Test was performed

because the data lacked a normal distribution. Within each Information Representation

Type, the SWAT-Recall responses were compared between all levels of Task Load (e.g.,

Low vs. Medium, Medium vs. High). For Nickname, perceived effort at Medium Task

Load was significantly higher than at Low Task Load ($z = -3.821$, $p < .001$). Likewise,

Nickname perceived effort at High Task Load as significantly higher than at Low Task

Load ($z = -4.05$, $p < .001$). For Non-Face, perceived effort at Medium Task Load was

significantly higher than at Low Task Load ($z = -3.025$, $p < .002$). Also, for Non-Face,

perceived effort at High Task Load was significantly higher than at Medium Task Load (z

= -2.149, $p < .032$). Also, for Non-Face, perceived effort at High Task Load was significantly higher than at Low Task Load (.$z = -3.833$, $p < .001$).

Table 11 Subjective Task Load (SWAT-Recall) means and standard deviations

Information Representation Type	Low	Medium	High
Nickname	9.29(5.83)	14.03(5.63)	14.38(5.59)
Non-Face	9.62(5.78)	12.65(6.27)	13.85(4.76)
Non-Face & Nickname	11.65(6.38)	12.88(5.59)	12.29(5.47)
Face	10.59(6.29)	12.71(5.96)	14.03(5.13)
Face & Nickname	11.47(6.04)	13(5.24)	12.91(5.15)

Additionally, the Information Representation types were compared to one another at each level of Task Load (e.g., Nickname at Low Task Load was compared to Face at Low Task Load, Non-Face at Low Task Load was compared to Non-Face & Nickname at Low Task Load). No significant differences were found between Information Representation Types at Medium Task Load. However, there were significant differences at both Low and High Task Loads across specific Information Representation Types.

For Non-Face & Nickname vs. Nickname (Table 12), at Low Task Load perceived effort for Non-Face & Nickname was significantly higher than Nickname ($z = -2.131$, $p < .033$). But at High Task Load, this effect reversed and perceived effort for Nickname was significantly higher than Non-Face & Nickname ($z = -2.945$, $p < .003$).

Table 12 Subjective Task Load (SWAT-Recall) means for Non-Face & Nickname vs. Nickname

Information Representation Type	Low	High
Nickname	9.29(5.83)	14.38(5.59)
Non-Face & Nickname	11.65(6.38)	12.29(5.47)

For Face & Nickname vs. Nickname (Table 13), at Low Task Load perceived effort for Face & Nickname was significantly higher than Nickname ($z = -2.569$, $p < .01$). But at

High Task Load, this effect reversed and perceived effort for Nickname was significantly

higher than Face & Nickname ($z = -2.489, p < .013$).

Table 13 Subjective Task Load (SWAT-Recall) means for Face & Nickname vs. Nickname

Information Representation Type	Low	High
Nickname	9.29(5.83)	14.38(5.59)
Face & Nickname	11.47(6.04)	12.91(5.15)

For Non-Face & Nickname vs. Non-Face (Table 14), at Low Task Load perceived effort

for Non-Face & Nickname was significantly higher than Non-Face ($z = -2.081, p < .037$).

But at High Task Load, this effect reversed and perceived effort for Non-Face was

significantly higher than Non-Face & Nickname ($z = -2.039, p < .041$).

Table 14 Subjective Task Load (SWAT-Recall) means for Non-Face & Nickname vs. Non-Face

Information Representation Type	Low	High
Non-Face	9.62(5.78)	13.85(4.76)
Non-Face & Nickname	11.65(6.38)	12.29(5.47)

Each of these sets of effects indicates an interaction between level of Task Load and the

Information Representation Type for the mean SWAT-Recall query responses.

CHAPTER FIVE: GENERAL RESULTS AND CONCLUSION

In this section, the study's findings are discussed in terms of the stated hypotheses.

Potential theoretical and practical implications are presented as well as suggestions for

future research.

<u>Hypotheses and Related Findings</u>

Hypothesis 1 predicted a main effect for Information Presentation Types across all levels

of Task Load with Face & Nickname outperforming all other types. The data did not

show a significant main effect for Information Representation Type. However, the

significant main effect for Task Load and associated mean recall scores do indicate that

the Task Load manipulation worked as intended. Furthermore, a significant interaction

effect did occur between each of the Task Load levels and the Information

Representation Types. Simple main effects revealed a changing pattern across the

different levels of Task Load. In the Low Task Load condition, Face significantly

outperformed Non-Face and Face & Nickname. Also, in the Low Task Load condition,

Nickname significantly outperformed Non-Face. At the Medium Task Load level, Face &

Nickname significantly outperformed Nickname, Non-Face & Nickname and Face. At

the High Task Load level, both Non-Face and Non-Face & Nickname significantly

outperformed Nickname and Face. The contrasts for the Face Information Representation

Types (i.e., Face vs. Face & Nickname) show that when comparing a Face to another

Face that has a nickname associated with it, the the presence of a nickname had a

significant impact on the performance difference for both Medium and High Task Load

levels. Interestingly, the contrasts for the Non-Face Information Representation Types

(i.e., Non-Face vs. Non-Face & Nickname) reveal that the presence of a nickname did not

yield any benefit to performance across all Task Load levels. The mean recall scores were virtually identical. This could mean that the Non-Face (i.e., fruit) images provided an effect powerful enough that nicknames were not required, and nicknames did not provide any form of interference with this effect. In the case of Non-Face & Nickname vs. Nickname, the contrasts indicate that at Medium and High Task Load levels, the presence of a Non-Face offered a significant performance difference and performance increase, particularly at High Task Load. The same effect holds true for Face & Nickname vs. Nickname. At higher Task Load levels, the addition of a Face to the Nickname benefits performance.

Hypothesis 2 predicted a main effect for response times for each Information Representation Type across all levels of Task Load. The results were not significant.

Hypothesis 3 predicted a main effect for perceived mental effort for each Information Representation Type across all levels of Task Load. The first two subjective Task Load queries focused on the task itself (SWAT-Effort and SWAT-Time), while the third focused on the effort associated with answering the post-trial recall query (SWAT-Recall). The results for SWAT-Effort indicate that, across all levels of Task Load, the trials with Non-Face required less effort than Face or Nickname. The results for SWAT-Time indicate that, across all levels of Task Load, the tasks with Non-Face produced less perceived time pressure than both Non-Face & Nickname and Face. Finally, the results for SWAT-Recall indicate that at high levels of Task Load, participants exerted more effort answering the Recall query in the Nickname conditions versus either the Non-Face

& Nickname or Face & Nickname conditions. This is somewhat consistent with the significantly lower Recall scores for Nickname vs. Non-Face & Nickname at High Task Load. However, participants still felt it was harder to answer the recall query in the Non-Face & Nickname condition versus the Non-Face condition. Since there was no significant difference between Recall scores for Non-Face & Nickname vs. Non-Face conditions, these results may indicate that simply using a Non-Face is both operationally effective and easier for a user when Task Load is high. All things considered, it seems this hypothesis was partially upheld in that participants perceived less mental workload when an image was involved versus the nickname alone, but the Non-Face condition proved to induce less mental workload than the Face condition.

<div align="center">Theoretical Implications</div>

These findings indicate that human faces, when coupled with a nickname, may serve to function as a kind of mnemonic device with respect to recall performance. If this is the case, then it is possible that face images used in this manner may impose less load on attentional resources when used to help encode information and when used as a retrieval cue. It appears that both Face & Nickname and Non-Face & Nickname as a group had generally higher mean recall scores than Face, Non-Face or Nickname across Task Load levels. It is likely that dual encoding (i.e., the combination of image and semantic label) played a role in delivering higher recall performance. Therefore, the multi- vs. one-dimensional nature of the Information Representation Type may be a key driver of the recall performance differences. The use of an appropriate dual-task paradigm may further clarify the exact nature of these effects with respect to which dimensions may or may not process through working memory more efficiently (e.g., Non-Face vs. Face).

Practical Implications

At each level of Task Load, there were significant differences between Information

Representation Types. These differences manifested in three different ways: how images

compared to nicknames, how the combination of nickname and image compared to

nicknames and how different image/nickname combinations compare to one another. For

example, at Low Task Load, Nickname significantly outperformed Non-Face, but that

was the only time a Nickname outperformed another Information Representation Type at

any level of Task Load. These results may indicate that, from a design perspective,

Nicknames may not be a good practical choice over an image such as a face or generally

recognizable object such as a piece of fruit, in terms of accuracy at higher levels of task

load. At both the Medium and High Task Load levels, image and nickname combinations

significantly outperformed nicknames. These may results indicate that, from a design

perspective, an image and nickname combination may offer significant performance

advantage, in terms of accuracy, over a nickname alone at high levels of task load. The

Face & Nickname combination significantly outperformed Non-Face & Nickname in the

Medium Task Load condition, but there was no significant difference between them at

the High Task Load condition. So, it is hard to say whether or not Face & Nickname

would truly outperform Non-Face & Nickname in a practical sense. However, in a

context where task load peaks only periodically, Face & Nickname may offer a

performance gain. These results may indicate that the Face alone would not show any

practical performance benefit at higher levels of Task Load. At each level of Task Load

there was no significant difference between Non-Face and Non-Face & Nickname, so it

appears that the addition of a Nickname to the Non-Face image may have had little

impact on performance. This could be an important finding to a designer dealing with limited screen space on a mobile device. These findings could be useful in various technical applications across a variety of fields. As educational technology moves classrooms to online virtual environments, the software applications used to facilitate teaching has borrowed from social media applications in that students and teachers establish online personas to communicate wit one another. It is possible that these findings could lead to use in these environments to reinforce lesson content or to help teachers remember specific aspects of key information associated with each one of their students during real time online classes. The medical emergency response domain may make use of these findings as well to help nurses or other key medical personnel to remember specific types of critical information associated with a group of triaged patients. For example, an augmented reality device such as Google Glass could host software that implements the aforementioned design principles in such a way that specific reminders on the display are coupled with faces to stimulate recall. Such a cue can be used in place of text to reduce screen clutter—a critical issue for devices such as Google Glass that can be obtrusive with respect to the user's field of view. In a triage environment, medical personnel could potentially use a device like this to rapidly assess, triage and treat a select group of patients.

<u>Suggested Future Research</u>

As stated earlier, the Face & Nickname combination significantly outperformed Non-Face & Nickname in the Medium Task Load condition, but there was no significant difference between them at the High Task Load condition. This effect should be studied further to see if one or the other could be adjusted to yield a more discernable

performance result. The lack of significant differences between the Non-Face and

Nickname and the Non-Face conditions is also remarkable. Especially in light of the

significant performance increase provided by combining a Nickname with a Face in the

Medium Task Load condition. It would be helpful to see if combining a Nickname with

specific image types is only feasible with specific image types. Further study is required

to address the myriad of aspects that may lend mnemonic properties to a human face

image, these include, but are not limited to: complexion, clothing, hair style/color, age,

gender, attractiveness, facial expression, profile vs. portrait, resolution, dimensions,

celebrity photos, combining photos with contextually relevant backgrounds (e.g., fire).

There may also be a nickname nomenclature and image (face or non-face) pairing that

may result in a stronger performance boost. It would also beneficial to learn if coupling

the face image with an alphanumeric label that is created by the user under similar

conditions used in this experiment would create a significant performance boost as well.

Performance differences between genders could be investigated as well. For example,

females have been found to exhibit better face recognition performance (Sommer,

Hildebrandt, Kunina-Habenicht, Schact & Wilhelm, 2013; Lewin & Herlitz, 2002) than

males. Investigations could include varying the gender of the face images across the

participant genders to see if the performance significantly changes if the gender of the

participant matches or differs from the gender of the face images they use. Time series

data was not recorded for this experiment. However, given the nature of the software

platform that was implemented here, it would be possible to record such data to

determine the exact nature of the participant's performance during the task. These data

may shed further light on how the participant behaved when manipulating each Information Representation Type.

Conclusion

In rapidly unfolding situations such as natural disasters, social media has begun to play a more critical role. In this context, social media platforms such as Twitter have been used both as a conduit for victims to reach authorities and as an aggregation tool for individuals and organizations to build an understanding of what is happening. Given the rapid and ongoing proliferation of social media, the sheer volume of both relevant and irrelevant information is growing. One approach to helping users sort through and organize such information is to build more computational power in the form of sophisticated statistical models to filter information down to it's most essential components and present them to the user. Another approach is to focus on how the information will be presented to facilitate easier processing on the part of the user. Since the brain appears to possess a particular facility for processing faces, it seems reasonable to assume that a face could actually serve the same purpose as an icon. One reason why this avenue might not have been pursued in the area of human computer interaction in the past is that computers had not yet reached a level of sophistication where presenting high resolution images of human faces as part of a user interface could be achieved at relatively low cost.

Most social media platforms operate on a paradigm whereby individual persona's are a major part of meaningful communication. Though the hallmark of the proposed design paradigm would be to use face images, it does not necessarily mean that the face images

from a specific persona would be used. Rather, generic face images could be assigned to groups of related information to help a user sort, organize and remember how that information fits into a bigger picture. As a first step towards determining the viability of such a paradigm, the purpose of this study was to determine if face images would offer performance increases in terms of accuracy and speed for recalling information associated with specific events presented in a Twitter-style format. Face images combined with nicknames were compared to other information representation types including: 1) nicknames, 2) non-face images, 3) non-face images combined with nicknames and 4) face images. The comparisons were made across three different levels of task load. The findings indicate that Face & Nickname combinations may have offered a significant performance increase in terms of information recall accuracy at medium task load. This performance increase occurred over 1) nicknames, 2) non-face images & nicknames and 3) face images. However, the face & nickname combination did not significantly outperform any of the other information representation types at the other levels of task load. Response times were recorded, but did not significantly vary between conditions for any of the information representation types. Since a familiar object like fruit was used as a non-face image comparison, it is possible that it's familiarity and relative simplicity as compared to a face provided a mnemonic advantage.

Further research should be conducted with respect to specific types of face images to determine what visual characteristics make them more or less suited for use as part of an overall software design paradigm. Likewise, different combinations of alphanumeric label types and face types should be explored to see if there are optimum combinations.

Furthermore, since faces have fundamental structural components (e.g., geons) that imbue them with specific pattern characteristics that are recognizable as faces, it is possible to create icons that are "face-like" in their unique visual features. Such "face-like" icons could give way to a kind of visual representation that blends both unique recognizability and general categorical assignment that could be easily processed through working memory, thus facilitating faster performance in information management tasks. It is possible that such "face-like" icons may forego processing in the phonological loop and could thus be used in environments where multi-tasking is required both visual and audio load is heavy. In that context, "face-like" icons may make it easier for operators to more easily listen to and converse with one another while managing and processing visual information on computer screens in front of them.

APPENDIX A: EXPLANATION OF RESEARCH

EXPLANATION OF RESEARCH

Title of Project: A novel interface paradigm to enhance situation awareness for crisis response operations

Principal Investigator: Anthony Costello, MS

Faculty Supervisor: Waldemar Karwowski, PhD

You are being invited to take part in a research study. Whether you take part is up to you.

The purpose of this study is to determine how different types of user interfaces can affect user's situation awareness in a crisis response task. It allows the Principal Investigator to partially fulfill the requirements of his doctoral program.

Following an informal briefing about the experiment, you will be asked if you are comfortable to proceed. If you are comfortable, you will be asked to: 1) fill out a demographic questionnaire 2) take a spatial ability test 3) perform an information management task as part of fictional scenario involving several emergencies in metropolitan area 4) provide feedback on the mental demands of the task. You will perform the management task first with one type of software and then with another type. The task consists of organizing incoming emails for the purpose of staying current on the events taking place (for example, the status of fireman responding to a fire). The system will periodically ask you scenario-related questions about the latest information regarding a given event via pop-up survey questions.

Volunteer participation in this research project will take place in the UCF College of Engineering's Media Interface and Network Design Laboratory located in Room 311 in Engineering II. We expect that you will be in this research study for one session lasting approximately 2 hours.

Compensation for your participation will be $20 for the two-hour session ($10.00 per hour). If you complete any part of the experiment, you will receive compensation for the time you have spent in the experiment. If you choose not to participate, you may notify your instructor and ask for an alternative assignment of equal effort for equal credit. There will be no penalty.

You must be 18 years of age or older to take part in this research study.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, or think the research has hurt you, contact: Anthony

Costello, Graduate Student, Department of Industrial Engineering and Management Systems, College of Engineering and Computer Science, (407) 312-9458 or by email at mistercostello@msncom or Dr. Waldemar Karwowski, Faculty Supervisor, Department of Industrial Engineering and Management Systems at (407) 823-5759 or by email at wkar@ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901.

APPENDIX B: DEMOGRAPHIC QUESTIONNAIRE

Please check the appropriate areas to the left of your answers.

1. Please check the age bracket that you fall under

_____ 18 – 25 yrs. old

_____ 26 – 40 yrs. old

_____ 41 – 55 yrs. old

_____ Over 55 yrs. old

2. What is your gender?

_____ Male

_____ Female

3. Please check any of the following physical characteristics that apply to you.

_____ Left Handedness (the property of using the left hand more than the right hand)

_____ Right Handedness (the property of using the Right hand more than the left hand)

_____ Ambidextrous (the property of using one hand no more than the other)

_____ Color blind in any way (if yes please describe below)

_____ Wear corrective lenses (reading glasses, bifocals, contact lenses, etc.)

4. What is your current job occupation?

_____ Clerical

_____ Engineering

_____ Film/Broadcasting

_____ Skilled Technician

_____ Quality Assurance

_____ Manager/Supervisor

_____ Health Services

_____ Telemarketing/Telecommunications

_____ Student (if student, please list major) Major: _

_____Circle year in school (if you are a student): FR   SO   JR   SR   GRAD

5. In general, how do you feel about working with computers?

_____ I don't like working with computers.

_____ I have no strong like or dislike for working with computers.

_____ I like working with computers.

6. What is your highest academic degree?

_____ no degrees

_____ High school degree

_____ Trade or vocational school degree (beyond the high school level)

_____ College degree (for example, B.A., B.S., Associate College degree)

_____ Graduate degree (for example, M.A., M.S., Ph.D., Ed.D., M. D., R. N.)

7. What is your native language?

_____ English (go to question 9)

_____ Spanish

_____ Other (please name)

8. If your native language is not English, how well do you read English (leave blank if

English is your native language)?

_____ Poorly (I have trouble reading documents in English.)

_____ Adequately (I read well enough to get around.)

_____ Fluently (I read almost as well as a native speaker.)

_____ Other (please describe)

9. Do you have any prior experience working in an Emergency Operations Center of any

kind?

_____ No

_____ Yes (Please explain briefly below)

10. How often do you use Twitter?

_____ I do not have a Twitter account

_____ I use it weekly

_____ I use it every two or three days

_____ I use it daily

APPENDIX C: POST-TRIAL SUBJECTIVE QUESTIONNAIRE

Participant ID: _____

Trial #:_____

Instructions: Please place an 'x' in a space on the scale that corresponds to the way you feel concerning each of the following questions.

1) Overall, how hard did you have to work (mentally) to accomplish your level of performance?

```
┌─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┐
└─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┘
  Low                  Medium              High
```

2) Overall, how much time pressure did you feel due to the rate at which the task occurred?

```
┌─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┐
└─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┘
  Low                  Medium              High
```

3) How hard did you have to work (mentally) to answer the previous question?

```
┌─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┬─┐
└─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┴─┘
  Low                  Medium              High
```

Notes: These queries were administered electronically. The first two queries were administered post-trial. The third query was administered <u>after</u> the post-trial recall query.

APPENDIX D: IRB APPROVAL LETTER

Approval of Human Research

From: **UCF Institutional Review Board #1**
            FWA00000351, IRB00001138

To: **Anthony Costello**

Date: **February 26, 2014**

Dear Researcher:

On 2/25/2014, the IRB approved the following minor modifications to human participant research until 07/18/2014 inclusive:

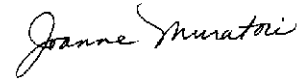|	
Type of Review:	IRB Addendum and Modification Request Form
Modification Type:	Study title has been changed from: A novel interface paradigm to enhance situation awareness for crisis response operations TO Mnemonics for Information Retrieval Under Workload. Another change is that the study will not be conducted in a physical lab, but on a secure, stand-alone website that the PI will maintain. Participants will be recruited using SONA system or via referral from professors on dissertation committee. Participants will be compensated $10 per hour for total of $20 if they complete the study activities. Only 40 participants, not 100 as initially approved, will be needed. Revised questionnaires have been uploaded in iRIS. A revised Informed Consent has been approved for use.
Project Title:	Mnemonics for Information Retrieval Under Workload
Investigator:	Anthony Costello
IRB Number:	SBE-12-08606
Funding Agency:	
Grant Title:	
Research ID:	N/a

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 07/18/2014, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

Page 1 of 2

73

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

IRB Coordinator

REFERENCES

Aaltonen, A., & Lehikoinen, J. (2006). Exploring augmented reality visualizations. In *Proceedings of the working conference on advanced visual interfaces* (pp. 453-456). Venezia, Italy: ACM.

Arnold, P., & Murray, C. (1998). Memory for faces and objects by deaf and hearing signers and hearing nonsigners. *Journal of psycholinguistic research*, *27* (4), 481-497.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The psychology of learning and motivation, 8*, 47-89.

Baddeley, A. D. (2002). Is working memory still working? *European psychologist, 7*(2), 85-97.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual review of psychology, 63*, 1-29.

Banbury, S. P., Macken, W. J., Tremblay, S. B., & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors, 43* (1), 12-29.

Bellezza, F. S. (1981). Mnemonic devices: Classification, characteristics, and criteria. *Review of Educational Research, 51* (2), 247-275.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review, 94* (2), 115.

Bruce, V. (2009). Remembering Faces. In J. R. Brockmole (Ed.), *The visual world in memory* (pp. 66-88). New York: Psychology Press.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology, 77* (3), 305-327.

Camacho, M. J., Steiner, B. A., & Berson, B. L. (1990). Icons vs. alphanumerics in pilot-vehicle interfaces. paper presented. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. *34* (1), 11-15.

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science, 30* (6), 433-464.

Czerwinski, M., Van Dantzich, M., Robertson, G., & Hoffman, H. (1999). The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3D. In *Proceedings of Interact*, *99*, 163-170.

Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology, 63* (2), 180.

Dumais, S. T., & Landauer, T. K. (1983). Using examples to describe categories. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* (pp. 112-115). New York, NY: ACM.

Ellis, H. D. (1975). Recognizing faces. *British Journal of Psychology, 66* (4), 409-426.

Ells, J. G., & Dewar, R. E. (1979). Rapid comprehension of verbal and symbolic traffic sign messages. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 21* (2), 161-168.

Fertig, S., Freeman, E., & Gelernter, D. (1996). "Finding and reminding" reconsidered. *ACM SIGCHI Bulletin, 28* (1), 66-69.

Fiore, S. M., Johnston, J., Van Duyne, L. (2004). Conceptualizing the Training Space: Constructing hierarchies to integrate time and space for distributed debriefings, In *Proceedings of the Human Factors and Ergonomics Society, 48,* 2562-2566.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of experimental psychology: General*, *108* (3), 356.

Hygge, S., Boman, E., & Enmarker, I. (2003). The effects of road traffic noise and meaningful irrelevant speech on different memory systems. *Scandinavian Journal of Psychology, 44* (1), 13-21.

Jensen, D. G. (1988). *Facial perception: A special case of visual perception?*, Rice University, Houston.

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21* (4), 1008.

Jones, D., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology, 44* (4), 645-669.

Jones, W., Phuwanartnurak, A. J., Gill, R., & Bruce, H. (2005). Don't take my folders away!: organizing personal information to get things done. In *CHI'05 extended abstracts on human factors in computing systems* (pp. 1505-1508). New York, NY: ACM.

Jones, W. P., & Dumais, S. T. (1986). The spatial metaphor for user interfaces: experimental tests of reference by location versus name. *ACM Transactions on*

Information Systems (TOIS), 4 (1), 42-63. New York, NY: ACM.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence, 73* (1), 31-68.

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 133* (3), 355.

Kleiman, G. M. (1975). Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior, 14* (4), 323-339.

Lansdale, M. W. (1988a). The psychology of personal information management. *Applied ergonomics, 19* (1), 55-66.

Lansdale, M. W. (1988b). On the memorability of icons in an information retrieval task. *Behaviour & Information Technology, 7* (2), 131-151.

Lansdale, M. W. (1991). Remembering about documents: memory for appearance, format, and location. *Ergonomics, 34* (8), 1161-1178.

Lansdale, M. W., Jones, M. R., & Jones, M. A. (1989). Visual search in iconic and verbal interfaces. *Contemporary Ergonomics*, 422-429.

Lansdale, M. W., Simpson, M., & Stroud, T. R. M. (1990). A comparison of words and icons as external memory aids in an information retrieval task. *Behaviour & Information Technology, 9* (2), 111-131.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition--Women's faces make the difference. *Brain and cognition, 50* (1), 121-128.

Logie, R. H. (1995). *Visuo-Spatial Working Memory*. Hillsdale: LEA.

Luximon, A. & Goonetilleke (2001). Simplified subjective workload assessment technique. *Ergonomics*, *44* (3), 229-243.

Malone, T. W. (1983). How do people organize their desks?: Implications for the design of office information systems. *ACM Transactions on Information Systems (TOIS), 1* (1), 99-112.

Meshkati, N., Hancock, P. A., Rahimi, M., & Dawes, S. M. (1995). Techniques in mental workload assessment. In P.A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: North-Hollad), 251-268.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review, 63* (2), 81.

Moody, C. (2012). *Meet FDNY's one-woman Twitter response team guiding New Yorkers*

through storm. Retrieved December 29, 2012, from
http://news.yahoo.com/blogs/ticket/meet-fdny-one-woman-twitter-response-team-guiding-141143449.html

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

Norman, D. (2002). *The Design Of Everyday Things*.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3* (4), 406.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., & Van Dantzich, M. (1998). Data mountain: using spatial memory for document management. In *Proceedings of the 11th annual ACM symposium on User interface software and technology* (pp. 153-162). San Francisco, CA: ACM.

Sanders, M. S., & McCormick, E. J. (1987). Human Factors In Engineering And Design (7th ed.): McGraw-Hill.

Saul, M. H. (2012). *Bloomberg: 911 System Swamped by Calls*. Retrieved December 29, 2012, from http://blogs.wsj.com/metropolis/2012/10/29/bloomberg-911-system-swamped-by-calls/

Schacter, D.L. (1999). The Seven Sins of Memory: Insights from psychology and neuroscience. *American Psychologist*, *54* (3), 182-203.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84* (1), 1-66.

Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using twitter to explore the ecologies of two climate change protests. *The Communication Review, 14* (3), 197-215.

Shiffrin, R.M. & Nosofsky, R.M. (1994). Seven plus or minus two: a commentary on capacity limitations. *Psychological review*, *101* (2), 357-361.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review, 84* (2), 127-190.

Sommer, W., Hildebrandt, A., Kunina-Habenicht, O., Schacht, A., & Wilhelm, O. (2013). Sex differences in face cognition. *Acta psychologica, 142* (1), 62-73.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46* (2), 225-245.

Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology, 87* (1), 116-124.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review, 80* (5), 352-373.

Visualscope. (2013). *What is the difference between Twitter and Facebook?* Retrieved December 29, 2012, from http://www.visualscope.com/twitfb.html

Whittaker, S., Matthews, T., Cerruti, J., Badenes, H., & Tang, J. (2011). *Am I wasting my time organizing email?: a study of email refinding.* In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3449-3458). Vancouver, BC: ACM.

Wickens, C. D., Hollands, J.G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Wiseman, S., & Neisser, U. (1974). Perceptual organization as a determinant of visual recognition memory. *The American Journal of Psychology*, *87* (4), 675-681.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81* (1), 141.

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Access to identity-specific semantic codes from familiar faces. *The Quarterly Journal of Experimental Psychology, 38* (2), 271-295.