

Quantifying episodic memories from real-world experience

Alyssa Marconi

Senior Thesis

Department of Biology
Emmanuel College
&

Kreiman Laboratory
Boston Children's Hospital, Harvard Medical School
Boston, MA

Principal Investigator: Dr. Gabriel Kreiman, Ph.D.
Advisor: Dr. Jason Kuehner, Ph.D.

April 27th, 2017

Abstract

Quantifying episodic memories from real-world experience

Alyssa Marconi¹ * Gabriel Kreiman²

¹Emmanuel College, Boston MA 02115 ²Children's Hospital, Harvard Medical School, Boston MA 02115, USA

Emmanuel College, Boston MA

Episodic memories form the fundamental fabric of who we are. These memories are full of content and malleable due to the number of inputs impinging upon our senses in our everyday lives. Despite a plethora of laboratory studies about memories for individual words, names, or pictures, little is known about what governs the formation of episodic memories in real life. In this work, we studied real life episodic memories by recording ground truth video and eye movement information during a one hour segment in the life of our subjects and subsequently evaluating their ability to recognize those events. Ten subjects completed an encoding task which took place in two versions: an outdoor and indoor route. Memory recognition tasks were completed at 24 hours up to 3 months post-encoding and performance was consistent across subjects in both versions with an accuracy of up to 62 +/- 7.0%. Previous work used a similar methodology combined with machine learning approaches to predict episodic memory formation in movies, but those results remain inconclusive since movies are artificial stimuli and do not accurately represent real life events. Our study found that real life episodic memories are less memorable than movies likely due to the large amounts of information we filter out in non-laboratory settings. The results allow us to understand episodic memory formation in real life to elucidate the subjective filtering which leads to how we learn and consolidate long-term memories. To evaluate the content properties that influence long-term memory formation, annotations were designed to see what aspects enhance memorability in the real world. These allow us to report what content properties correlate with higher performance in memory recognition. To the best of our knowledge, this work constitutes the first quantitative approach to directly measure memory formation with ground truth data in real life scenarios.

Table of Contents

Title Page	1
Abstract	2
Table of Contents	3
List of Figures and tables	4
Introduction	5-18
Episodic Memory Formation and Retrieval from Real-World Experience	5-6
Alzheimer’s Disease and Episodic Memory	7-8
Learning disabilities and Episodic Memory	8
The stages of the human memory process	8-9
Learning and Memory	9-10
Patient H.M. and the evidence for different Memory Systems in humans	10-13
In-lab Episodic Memory Studies from previous literature	14
Memories in Real-Life	14-15
Quantitatively characterizing efficacy of long-term memorability	15-16
Spatiotemporal memory formation in movies vs. real life experience	17-18
Materials and Methods	19-28
Results	29-43
Discussion	44-52
Conclusions	53
Acknowledgements	54
References	55-56

List of Figures

[1] Real-world eye-tracking apparatus for encoding (Part I) and footage for the memory recognition task (Part II)	6
[2] The three stages of the human memory process	9
[3] Overview of the different memory systems in humans	12
[4] Performance on the memory recognition test was consistent at both time points and roughly equal in the categorization of scenes vs. faces	32
[5] Performance on the memory recognition test was retained across subjects on average 3-4 months post-encoding	32
[6] Receiver Operating Characteristic (ROC) Analysis for 30 hours vs. 3-4 months post-encoding showed a decrease in both the False Positive Rate (FPR) and True Positive Rate (TPR).	34
[7] Average Shot Content Properties correlated with performance across the subject pool (A-J). Performance for the memory recognition test that corresponds to if the content was present in the shot or not (mean +/- SEM).	37
[8] Performance in the MFA variant was consistent with version one and had no statistically significant difference in scenes vs. faces	39
[9] Receiver Operating Characteristic (ROC) Analysis for the MFA variant	41
[10] Shot Content Properties correlated with performance (A-H). Performance for the memory recognition test that corresponds to if the content was present in the shot or not (mean +/- SEM).	43

List of Tables

[1] Summary of subjects and test conditions in the main experiment and MFA variant	31
--	----

Introduction

Quantifying episodic memories from real-world experience

Episodic Memory Formation and Retrieval from Real-World Experience

Memories are the fabric of who we are and include personal experiences that occurred a specific time or place in the past. Episodic memories, commonly referred to as autobiographical memories, are memories that bring you back to a specific time or place and include the who, what, when, where, and why knowledge (Tulving 2002). These types of memories allow us to travel back to the time of that event and recollect the specific context in which it happened. Episodic memories are types of declarative memory (fig. 3). To understand episodic memories, we must decipher the mechanisms involved in the acquisition, retention, and retrieval of knowledge (Gabrieli 1998). Although episodic memories are detailed, they can also be extremely malleable, and large amounts of the details within them are typically forgotten (Tang et al. 2016). Memory is malleable because it is shared, constructed, elaborated, and exchange with social interactions (Cohen 1996, p.9). The main research question being addressed in this study is quantifying what people remember from episodic memories *in real life*. In the field of neuroscience, the biological mechanism for the formation and retrieval of the specific contents of episodic memories is still not clear. Previous studies have addressed this question and identified cues but have focused on studying individual words or images, lacking fundamental components of episodic memories including realistic context, which this study aims to address.

One of the central challenges to studying memory formation in real life scenarios is the lack of ground truth data to evaluate the accuracy of memory recollections. To study the formation and retrieval of episodic memories in real life, we collected ground truth data from subjects during the first portion of the experiment known as “encoding” (fig. 1). This data from

the subjects' experience was used in a memory recognition test that took place 24 hours and again at 3 months post-encoding (fig. 1c and 1d). In the context of this study, ground truth data refers to data that we collected about specific sequences of events experienced by the test subjects as described below. In the case of our study, the ground truth data refers to footage obtained with a *GoPro* camera and *Eye-link* eye-tracking results collected when the subject performs the encoding portion of the experiment (fig. 1a), either in version 1 (outdoors), or in version 2 which took place indoors (fig. 1e and 1f). The extent to what subjects recall from their episodic experience depends on a multitude of factors including: age, time between encoding and the memory recognition test, context of the episodic memory, and similarity between the clips the subject actually did see (target clips) and control clips the subject did not see (foil clips). By evaluating a subject's memory of their episodic experience using ground truth data in natural conditions we can better understand how we collect and retrieve parts of our everyday lives. In this way, we can quantify what aspects of a memory people can retrieve.

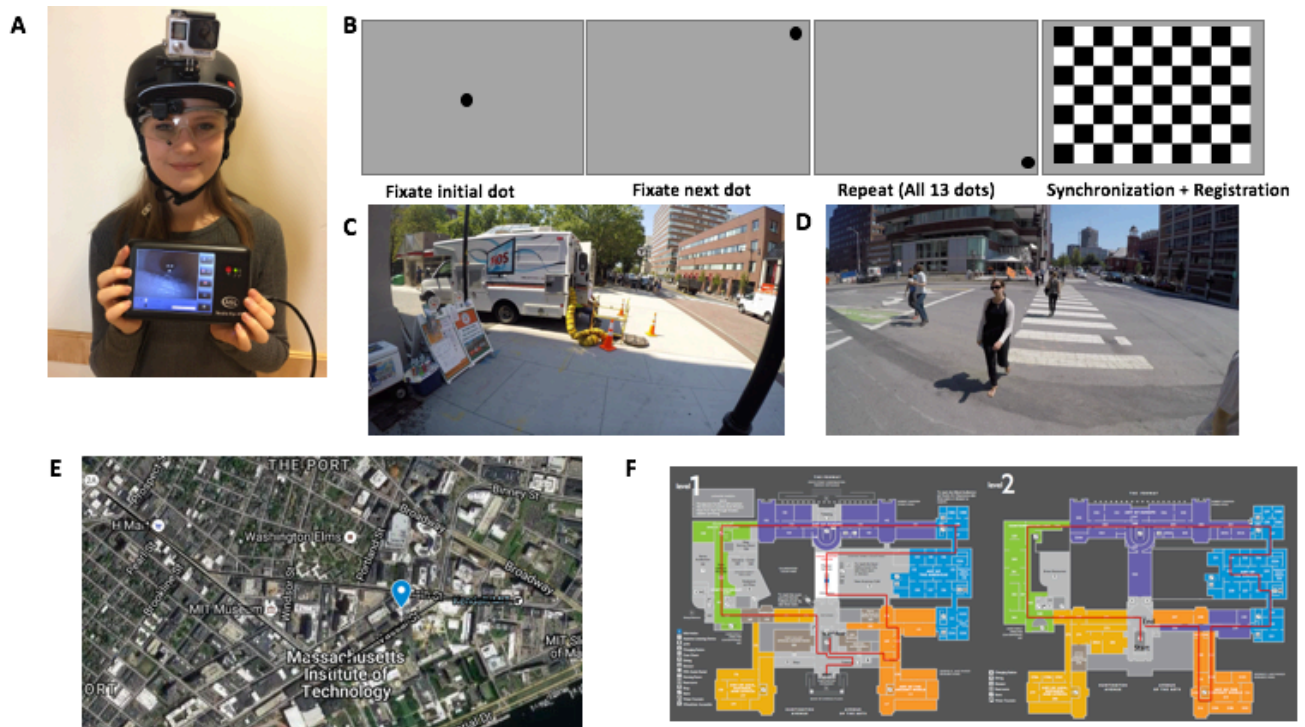


Figure 1: Real-world eye-tracking apparatus for encoding (Part I) and footage for the memory recognition task (Part II). (A) Subjects were fitted with a pair of ASL eye tracking glasses and a supplemental GoPro camera that enhanced the quality and field of vision (FOV) of the recorded video of the subject's visual environment which was mounted to a helmet. (B) Calibration (moving dot) and ASL-to-GoPro synchronization and registration (checkerboard pattern) were automated and standardized across participants using MATLAB psychophysics toolbox. (C + D) An overview of the memory recognition tasks subjects complete post-encoding including "scenes" or "faces." Subjects are shown an equal amount of "target" clips from their experience and "foil" clips those that are not from their experience and asked to choose if they remember the clip in a "yes" or "no" task where clips are pseudo-randomized and appear in one-second segments. (E +F) Route maps for the two versions of this experiment, outdoors (E) and indoors at the MFA (F). These routes were controlled and where subjects were instructed to walk during their encoding experience.

Alzheimer's Disease and Episodic Memory

By quantifying what people remember from an episodic memory through a range of characteristics, then we can learn and apply these findings to a number of disciplines within neuroscience as well as gain insight into how people learn. Almost every aspect of our daily lives is influenced by memory. Since memories are the internal mental records we keep, they allow us to access our personal past, which includes all of the skills we have acquired and facts we have learned. When a person is not able to form new memories or recall previous memories, life becomes extremely difficult. This is particularly seen in neurodegenerative disorders where the key cognitive decline includes memory loss such as Alzheimer's and Dementia. Dementia, which is a symptom of Alzheimer's, makes it a truly debilitating disease that is not only becoming widespread in the US but the world. The harsh reality is that as time progresses, the disease is only predicted to become more prevalent because people's life expectancies will increase due to other medical advances. It is currently estimated that the number of people affected with Alzheimer's Disease is 36 million and is expected to triple by 2050 (Huang 2012). The relation between Alzheimer's Disease and episodic memory is that they are both associated with the same brain area: the hippocampus within the medial temporal lobe. In Alzheimer's Disease, the

hippocampus is where the most loss of brain volume appears due to loss of neuronal process which leads to shrinkage (Huang 2012). This is seen notably in PET imaging scans where hypo-metabolism is observed within the hippocampus and can spread to other brain regions as the disease progresses (Huang 2012). In a healthy brain, there is activation of the hippocampus during the execution of memory tasks. However, in a brain of a person with AD, there is abnormal network activity within the neural networks which actively interferes with the processes involving: learning, memory and overall cognitive functioning (Huang 2012). Therefore, damage to the hippocampus is what drives AD and leads to episodic memory deficits.

Learning disabilities and Episodic Memory

Another important discipline this study can be applied to is helping those with learning disabilities. This is because learning and memory are dependent upon one another and are associated with the hippocampus. The hippocampus is responsible for integrating the information we learn and plays a major role in how we later retrieve this information (Smarr 2014). This can allow us to understand how people learn, which can aid in the approach to teaching those with learning disabilities. In the field of neuropsychology, the presence of interference that can occur while children learn will prevent proper memory recall (Straube 2012). This is because there are three stages involved in the human memory process and any disruption in these stages will negatively impact recall.

The stages of the human memory process

The three stages of the human memory process: encoding, storage, and retrieval, allow us to experience the memory, and ultimately we remember some amount of that particular memory.

The encoding stage of memory, which is what subjects have on the first portion of the experiment, is when information is sent to the brain where it becomes dissected into its

significant composing elements (Straube 2012) (fig. 2). The various brain processes ensemble the incoming stimuli and translates this information into a neural code (Straube 2012). In the storage portion of memory formation, the human brain filters, selects and retains part of this encoded data over time. Lastly, the retrieval process, which is studied in the second portion of the experiment during the memory recognition phase, involves the entry into the infinite world of stored information within a person’s brain where old information comes out of permanent storage (Straube 2012) (fig. 2). In this part of the experiment, retrieval, subjects are shown segments from their encoding experience and prompted to select either “yes” or “no” in response to if they remembered seeing that clip.

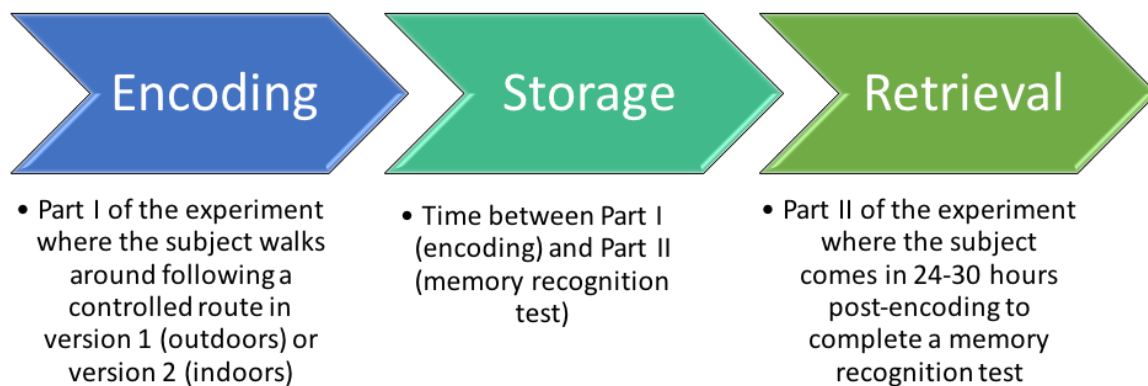


Figure 2. The three stages of the human memory process. *Encoding is the first stage of human memory and is the first part of the experiment where subjects walked around in Version 1 or 2 with the eye-tracking apparatus. Storage is the process where memories are sorted. Retrieval (Part II) is the last stage of human memory where specific experiences are either filtered or remembered from the encoding experience.*

Learning and Memory

When defining what "learning" is in children and adults, it simplifies to the idea of having the capability of the brain to modify information already stored in a memory based on a new input or experience (Straube 2012). Essentially, the first step of memory is learning, which means that our sensory systems collect the information we experience and send this information

to the brain. In a normal day, humans' sensory systems are constantly perceiving vast amounts of information each second (Straube 2012). However, only a small fraction of this is retained, and it has been found in both versions of this study as humans, we retain very little of our experience because it is impossible to remember every aspect of our daily lives. Our minds are constantly being bombarded with external information, and we make every effort to remember such a small amount of it through filtering during this memory process (Tang et al. 2016). The relationship between learning and memory is essential to discuss because memory cannot occur without learning (Straube 2012). Therefore, through quantifying what aspects of episodic memory make an experience more memorable, we can find what stimuli enhance the learning process in children.

Patient H.M. and the evidence for different Memory Systems in humans

A lot is known about the formation and retrieval of episodic memories and exactly what part of the brain is associated with memory due to the patient, Henry Molasin, known as patient H.M. This patient had epilepsy and as a result underwent neurosurgery that included bilateral resection of his hippocampus that encompassed two-thirds of the total volume (Squire 2004). This was because his seizures were localized to the medial temporal lobe, where the hippocampus is located. The behavioral and cognitive abnormalities observed in the patient were that he was able to function and complete tasks associated with short-term and procedural memory but could not consolidate new memories. This finding suggested the absence in the retrieval of long-term episodic memory (Squire 2004). Patient H.M.'s contribution to science was significant because it allowed scientists to understand the organization of human memory and the multiple memory systems it possesses (Squire 2004). The unexpected discovery that there were several types of memory was found when patient H.M. had his ability to acquire and

test his visuomotor skills since he had global memory impairment in his hippocampus. Patient H.M. was shown a five-pointed star and asked to trace its outline with a pencil. It was evident that he was acquired the mirror-drawing skill because after about ten trials he was able to have retention of the task for up to three days after (Squire 2004). However, at the end of testing the patient exhibited no recollection of having completed the task prior. This was a vital finding that paved the way for the basis of different types of memories. This specific type of memory tested: the memory of motor skills must have lied somewhere outside of the medial temporal lobe within the brain (Squire 2004). This was supported later by studies that found patients with amnesia were still able to perform skill-like abilities, similar to that of patient H.M. This made the groundwork for two main domains of memory being established: declarative memory and non-declarative, also known as procedural memory (Squire 2004). Declarative memory was the type of memory specifically absent in patient H.M, which is where episodic memory falls under and is present in the hippocampus of the medial temporal lobe. The other types of memory, known as non-declarative memories, are all other types of memory involved with skill and habit learning. These types of memories in procedural learning were still present in the patient H.M. patient which included the motor skill learning of drawing a star (Squire 2004). This finding was essential to learning the multiple memory systems in mammals.

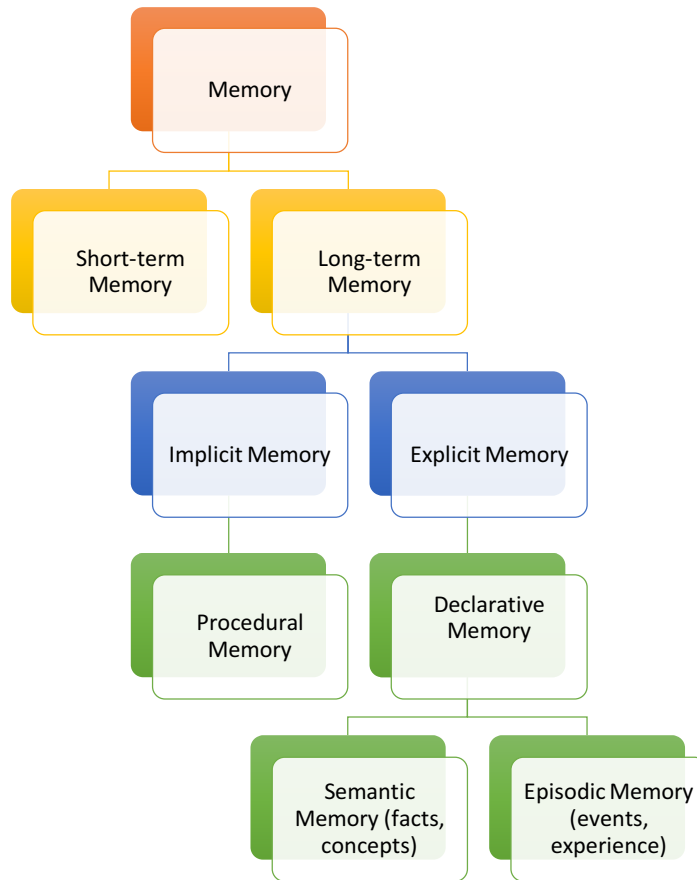


Figure 3. Overview of the different memory systems in humans. *The type of memory being studied within this experiment is episodic memories. They are a type of explicit long-term memory that are declarative, meaning it includes facts and events. Episodic specifically focuses on an event or an "episode."*

Patient H.M.'s memory impairment reflected the failure to convert transient memory into stable long-term memory. Similar to that of patient H.M., other patients with hippocampal resection known as patients, E.P. and G.P. had severe memory impairment (Squire 2004). In these two patient's cases, they also had issues with episodic memories (fig. 3). It was found that the patients had their autobiographical memory impaired from the recent past, but as for the remote past, they were intact (Squire 2004). This finding also supports the idea of long-term memory. The findings of all of these patients allow us to recognize the brain areas associated with specific types of memories. Since episodic memories are considered to be declarative, it includes the acquisition, retention, and retrieval of knowledge (Gabrieli 1998). Episodic

memories, which can be measured by direct or explicit tests of memory such as recognition, all refer to the prior episode of encoding that occurred (Gabrieli 1998).

Within my study, we are focusing on visual encoding and visual recognition within the episodic memories. As mentioned previously it is known that the hippocampus plays a role in episodic memory, which is located within the medial temporal lobe. However, other brain areas associated with memory play limited roles in episodic memories specifically (Gabrieli 1998). These structures include the amygdala and the inferior prefrontal gyrus. It has been found that the amygdala is associated with fear activation and is activated during the retrieval of autobiographical memories that specifically have personal emotional value (Gabrieli 1998). In short, the amygdala is associated with processing negatively salient stimuli (Gabrieli 1998). Additionally, within the inferior pre-frontal gyrus it has been found that left frontal activations in the anterior portion of this structure are enhanced when shown faces (Gabrieli 1998). This type of recognition within declarative memories is declined in those with degenerative or developmental disorders including Parkinson's Disease, Huntington's Disease, and Gilles de la Tourette's Syndrome (GTS) (Gabrieli 1998). This type of working memory, reasoning and strategic memory performance declines linearly across a life span due to a decrease in Dopaminergic functioning, about 5-10% per decade (Gabrieli 1998). In this study subjects' were between the ages of 18-22. Therefore, their episodic memory should be intact and wouldn't expect to be in a decline since they were all healthy controls.

In-lab Episodic Memory Studies from previous literature

Before this study there had been several within the field that focused specifically on the recollection of words, faces, objects, or scenes (Bahrick 1975, Rubin 1996, Brady 2008, Vogt 2007, Standing 1973, Castelhana 2010 & Andermane 2015). These studies included in-lab

memory recollections of words, pictures or faces. Although these studies have been interesting to study memory consolidation and recall, they lack spatial and temporal contexts that are critical to memories that are encoded and retrieved in real life (Tang et al. 2016). By testing a subject on specific aspects of their episodic experience using ground truth data, we can specifically see the formation and retrieval which gives more depth than simply showing subjects picture or words and having them later recall if they saw them. The studies from previous literature focused within the lab have an artificial context where the subject is solely focusing on the task of encoding and retrieval. When taking an episodic memory study out of the lab, subjects experience all aspects of everyday life and experience the spatial and temporal contexts needed to properly evaluate episodic memory formation and recall. This study allows for a better understanding of these memories, where as in-lab studies from previous literature found results that can be difficult to translate to real-world scenarios.

Memories in the Real-Life

In this way, it is important to discuss real life memory and the limits it possesses in relation to traditional laboratory research (TLR). One important aspect to note is that most studies in the field of memory are concerned with intentional memory meaning the subject has a certain exposure to an element they are being asked to remember that is controlled by the experimenter (Cohen 1996, p. 9). Most of this studies are mainly concerned with the limits of capacity the human brain has for remembering certain pictures including faces or scenes. The faces people are shown are typically isolated faces that are taken out of context with all additional information stripped away (Cohen 1996, p.108). When studying episodic memory in real life faces are not seen in isolation; we recognize people; not faces alone (Cohen 1996, p. 108). In the setting of my study, a sole face is not shown it includes the face of the person and

the clothes a person is wearing and the context in which the subject passed them. Both clothes and context can act as a reliable cue for remembering that does not constitute a face alone which is typically seen in the isolated laboratory settings (Cohen 1996, p. 108). It is also important to note that in real life we rarely recognize isolated faces alone, we remember a person by everything that encompasses them within that particular moment we have exposure to them. Additionally, in realistic situations as in my study, learning to remember a face during the subjects encoding experience as an experimenter I had no control over the amount of exposure or the degree of attention paid to the peoples faces by the subjects (Cohen 1996, p. 108). These are all factors of studying memory recognition in real-life which are not taken into account in previous laboratory studies but was addressed within my study.

Quantitatively characterizing efficacy of long-term memorability

Efforts to study episodic memories in non-lab settings in previous studies involved several challenges including the establishment of ground truth data, reproducibility of results, appropriate controls and amount of practice or exposure to other variables (Moscovitch 2006, Bahrick 1975 & Rubin 1996). To tackle the challenges posed within the field, a previous study in the lab, *Predicting episodic memory formation for movie events* looked at the formation of episodic memories under natural conditions where movies were used as stimuli (Tang et al. 2016). This study has allowed subjects to form vivid and detailed memories of movie events as assessed by cued recall, recognition and metamemory estimates (Tang et al. 2016). Within this study, functional algorithms were able to explain what dynamic events will be remembered. This allowed for the production of quantitative predictions about how subjects both learn and form new memories under similar to real world conditions. A quantitative analysis of the relationship between event content and the filtering of events leading to the formation of memories under

dynamic, close to real world scenarios was characterized, and the robustness of long-term memory was investigated (Tang et al. 2016). This study was implemented by quantitatively characterizing the content variables that dictate the retrieval in episodic memories with retention times of up to one-year post-encoding through the combination of psychophysics measurements and the extensive range of content annotations (Tang et al. 2016). The results of this study showed that the subjects scored remarkably well on recognition memory tests, well above chance (50%) at 85.6 +/- 5.3% and below ceiling level (100%). The high performance demonstrated in these results showed high memorability in situations that are close to real life scenarios where stimuli are embedded in a spatiotemporal context that is complex.

This work extended the studies of just single items that had been looked at in the past as mentioned previously (Bahrick 1975, Rubin 1996, Brady 2008, Vogt 2007, Standing 1973, Castelhana 2010 & Andermane 2015). The nature of this study regarding consistency, accuracy, and malleability of memory supported the findings of similar studies within the field including brief stories and single items (Bahrick 1975, Rubin 1996, Brady 2008, Vogt 2007, Standing 1973, Castelhana 2010 & Andermane 2015). However, this study extended previous knowledge through the incorporation of spatiotemporal episodic sequences and establishes long-term memorability as a variable that is part of the encoding process of memory formation. Several of the annotated properties of each shot showed correlation with the performance of subjects. For example, subjects performed better, meaning getting a higher percentage correct, when a shot included action versus a shot without action.

Spatiotemporal memory formation in movies vs. real life experience

The study, *Predicting Episodic Memory Formation for Movie Events*, previously done in the lab was a promising start to understanding episodic memory formation and retrieval in close

to real world scenarios. Consequently, it is important to note that although movies may be similar to real life scenarios, they are not real life. More events happen in a movie than what happen in a person's day. It is important to note that in movies characters and events are typically more distinct than those in real-life and are meant to be memorable to viewers. Therefore, it was vital to understand memory formation in actual real-world events using ground truth data as evidence. The research project I was involved in extends from the previous Tang et al. study, *Predicting Episodic Memory Formation for Movie Events*, by having subjects walk around a controlled route (version 1 or version 2) (fig. 1e and 1f) to study their formation of an episodic memory through consolidation and visual recall. Subjects are shown one-second segments from their experience that are annotated for certain variables, as general as “faces” vs. “scenes” (fig. 1c and 1d).

A more in-depth look into the subcategories of faces include variables such as gender, the number of people, movement of people, interaction of people, and distinctiveness of people in that segment. In cases known as “non-faces,” or referred to as scenes, we looked at the subcategories including movement vs. no movement, distinctive scenes vs. non-distinctive scenes, and seeing if people vs. no people. In this way, we were able to determine the shot content properties are correlated with performance in the memory recognition task by reporting what content in each shot corresponded to a greater memorability in subjects. Episodic memories constitute a minuscule fraction of what people remember in the real-world. Therefore, it is hypothesized that the specific contents relevant for episodic memory formation will include more distinctive events or faces people encounter during their episodic experience than factors which would be less distinctive. In this way real world, episodic memories can be quantified by what subjects are likely to remember from their experience, or what they are likely to forget. This

study was done in two versions: one outdoor version where subjects walked around the area of Cambridge by MIT and an indoor version which took place at the Museum of Fine Arts (MFA) in Boston, MA (fig. 1f). The first version of the study exemplified everyday, real-world experience where subjects walked around areas of Cambridge that included unfamiliar settings to make it a truly episodic experience and control for exposure to the area and people.

The second version in the MFA focused on the episodic retrieval of distinct artwork and people at the MFA. This study included eye tracking to get a better idea of how fixations contribute to the formation of episodic memories. From the current results of version one of the study, it has been seen that real life is much less memorable than movies with subjects scoring significantly lower than the Tang et al. study with an average performance of to 55.7+/- 3.7% (mean +/- SD) which is slightly above chance (50%) and well below ceiling (100%). The MFA study has shown similar results with an average performance across subjects of 62.3+/- 7.0% but a slightly higher average due to the more distinctness of the artwork within the episodic event than the outdoor version. This study will allow for advancement in the field of episodic memory research in the field by a real world spatiotemporal context that has not been studied previously in the field.

Materials and Methods

Subjects:

A total of 9 subjects participated in the main experiment (outdoor route) and 10 subjects in the MFA variant (indoor route). All of the subjects were college students between the ages of 18 and 22. All tests were performed with the subjects' consent and followed the protocols approved by the Institutional Review Board.

Part I: Encoding portion of Episodic Memory

Subjects were told that we were interested in assessing everyday, natural visual experience. They were not given a task or told to look at anything specific, just to walk along the assigned route. First, subjects were fitted with mobile eye tracker glasses and a *GoPro Hero 4* camera (fig. 1a) before the initial calibration. The experimenter then accompanied the subject around the lab and nearby hallways of the Brain and Cognitive Sciences Building to show them the beginning of the outdoor route. Subjects were instructed to follow a mapped 2.1-mile route that included the MIT area, to Central Square and back to Kendall Square following Massachusetts Avenue (~55 minutes) (fig. 1e). The experimenter accompanied the subject by walking behind them only to ensure that everything was properly recording and that they were following the proper route. The experimenter then brought the subject back into the lab to conclude the study. All subjects were ran at the similar time of day (+/- 30 minutes) under similar outdoor weather conditions to maximize the between-subject consistency of environmental conditions.

ASL Mobile Eye-tracking Apparatus to Assess Visual Experience:

A mounted Mobile Eye XG unit was fitted on the subject along with the *GoPro* mounted to a Helmet. The Applied Science Laboratory (ASL) Mobile Eye-XG Tracking Glasses (ASL

Eye Tracking, Bedford, MA) measure real-world gaze direction at 60 gazes per second. The ASL glasses utilizes two cameras: a scene camera and an eye camera. These two cameras work together to estimate the position of fixation of the visual world of the subject. The scene camera records the visual scene at 60 frames per second (fps) and sits on top of the rim of the glasses (fig. 1a). It was adjusted to align to the center of a person's field of view (FOV). The eye camera records an infrared (IR) image of the subject's right eye. It reflects off of a partially IR reflected coated lens and was adjusted to reflect three dots were centered onto the person's pupil. These two cameras work together to detect a subject's corneal reflection and pupil through three dots which were produced by an IR emitter. One specific dot of the three that was reflected onto the pupil was known as the "primary spot." The primary spot was the dot that was least likely to be obstructed (Peterson et al. 2016).

To improve the ASL's scene camera FOV, video quality, and resolution, a *GoPro Hero 4 Silver* (GoPro, San Mateo, CA) was used recording at 30 fps. The *GoPro* was mounted on a Giro bike helmet using a *GoPro* front helmet mount. The *GoPro* camera was mounted on the center of the helmet which was positioned 3.5 inches (y-direction) above the scene camera and .5 inches (x-direction) to the right (fig. 1a). The *GoPro* camera has a fish-eye distortion, for this reason, the fixations analyzed when the two cameras were synchronized includes mainly ones within the person's central region (Peterson et al. 2016).

Initial Calibration of the ASL Eye-tracker and GoPro Camera:

Once the *GoPro* and eye tracker were properly fitted to the subject, they completed a standardized calibration task using the Psychophysics Toolbox 3.0.10 (Brainard, 1997) written in the program MATLAB (Mathworks, Natick, MA) on a 13" MacBook Pro laptop (Apple, Cupertino, CA). Subjects were first asked to fixate on a centrally presented dot for about two

seconds that contains a white circular center (fig. 1b). Once the subject was confident that they fixated on that dot they were asked to press the spacebar. After this, the same dot reappeared randomly to one of the other 12 positions out of the total 13, that includes the initial dot. The dot positions were arranged in a 4 x 3 grid space on the screen. Once all 13 dots were fixated upon, the entire array of dots appeared. At this point, subjects were asked to look again at each dot starting at the upper left corner and moving across each row (fig. 1b).

After the encoding session, this calibration data was used to manually calibrate the ASL eye tracker using ASL's Eye XG software. In this process, the scene camera footage was viewed at 8 fps which was the slowest motion in the program. For each dot transition the subject looked at, the cursor was fixed on that dot to confirm that it was where the subject was looking. Once the subject's fixated on a new dot, the cursor was moved, and this was continued for the duration of the calibration. A computer mouse was used to manually select the location the subject fixated on. The ASL Eye XG software computes a function which maps the displacement vector from the eye camera to the dot locations of the scene camera for each of the 13 dots the subject fixated upon during the calibration routine (Peterson et al. 2016).

Fixation Detection of the ASL Eye-tracker:

The participant's gaze location was recorded in (x,y) coordinates that were relative to the scene camera image for the valid frame estimated by the ASL Eye XG software. Frames that were used for analysis did not include: blinks, or extreme external IR illuminations and were excluded from analysis.

Fixations were defined by the ASL software's algorithm as an event where there were six or greater consecutive samples that were within one degree of the sample group (Peterson et al. 2016).

Synchronization of the ASL Eye Tracker and GoPro:

To sync the video footage from the ASL eye tracker to the HD *GoPro* footage a 12x7 checkerboard pattern was presented on the monitor during initial calibration through the psychophysics toolbox using MATLAB (fig. 1b). An automated synchronization script through MATLAB searches for the first frame in the eye trackers scene camera where the checkerboard is detected. MATLAB detects this frame with the checkerboard for the *GoPro*. The two videos become synchronized by aligning the checkerboard onset times. From this, a projective linear transform matrix maps the 192 vertex points from the ASL to the *GoPro*'s coordinates. This matrix is used to map gaze coordinates for each frame and each fixation event from the ASL to the *GoPro* videos (Peterson et al. 2016).

Recalibration of Eye-tracker and GoPro Camera:

To validate the participants gaze coordinates throughout the encoding portion of their study, recalibrations were regularly performed every 5 minutes denoted by the subject's phone timers. In the MFA variant, recalibrations occurred every 10 minutes. During this time the subject held the 12x7 checkerboard at arm's length that was centered at eye level. They were instructed to fixate for two seconds each at the upper left labeled "1" upper right labeled "2" lower left, labeled "3," lower right labeled "4," and the to the center labeled "5." The subjects were then asked to reset their phone timers for another five minutes to keep recalibrating during the walk which was around ~50 minutes in total.

Part II Recognition and Memory Evaluation:

After part I of the experiment, *encoding*, all subject's came back 30 hours to the *Center for Life Sciences* building to complete the memory evaluation portion. Prior to the subject coming in, there was no mention about studying or testing memory within the study, but it can be

rationalized that subject inferred that memory was involved by the fact that they were asked to come back 30 hours after encoding to a different location than prior. Subjects were told they would be presented with 1,050-one second segments (except one subject saw 630 clips and had a slightly altered route) and have to select “yes” or “no” via a computer mouse if they remembered that clip. Target or foil shots were shown in pseudo-random order with equal probability, chance performance was considered 50% and ceiling was 100%.

Subjects were also asked to come back to complete an additional memory recognition test 3-4 months post-encoding the same way they were presented 30 hours after encoding. However, this time to allow for variability subjects were shown a different 500 foil clips mixed from another subjects experience than seen before. All of their target clips remained the same.

The clips were shown in 30 fps on a 21.5" iMac desktop (Apple, Cupertino, CA). Each one-second clip was shown in dimensions of 760x1352 with a frame rate of 30 fps (Tang et al. 2016).

Definition of Target and Foil Clips:

In each trial, subjects were presented with a target or foil shot that was randomized. Target shots were defined as shots the person saw during their episodic experience. This included 500 shots (excluding the participant that only saw 300), half of those (250) were defined as “faces” shots and half (250) were defined as “non-faces,” or “scenes.” Face clips included shots where a person's face was seen within the shot. Scene clips were defined as shots that did not have a persons face directly within the clip, however, could have included someone far away or people from the background. This was because since we cannot control for the people on the route, it was nearly impossible for no people to be within the shot for all of the scene clips.

The foil clips were taken when a control subject walked the same route, at the same time of day under the same weather conditions, therefore there were no major distinguishing factors that would make a participant think it was obvious one clips were there's or not. Foil shots were taken from all sections of the entire walk, as were the subjects target clips. This allowed for a natural comparison of targets and foils. Between subjects' foils were consistent for the first five subjects and were switched for the next five to account for weather changing and therefore the clothes of people on the streets changing. There was an equal amount of both target and foil shots shown that were randomly shown along with fifty shots that were repeated within the experiment to test for self-consistency of the subjects.

Main Experiment:

The performance of the subject was evaluation in one session that took place 30 hours post-encoding. Subjects were offered a monetary incentive only when coming back in to complete the memory evaluation after encoding. This included compensation for both parts.

To evaluate the self-consistency of subjects, which was unknown by the subjects, a small amount (0.05%) of the shots were repeated randomly during the test; this was 50 shots out of the total 1050 shown during memory evaluation. The repeat trials were equally distributed between the foil and target shots.

MFA Encoding Variant 1:

In this experiment variant subjects walked around the Museum of Fine Arts (MFA) in Boston, Massachusetts for the encoding portion of the experiment for about 45-50 minutes on an assigned route throughout both levels of the museum. Subjects were recruited to come in on weekdays between the hours of 12 pm and 2 pm to maintain consistency. The subjects were fitted with apparatus (fig. 1a), and calibrated (fig. 1b) at the Boston Children's Hospital Center

for Life Sciences Building and escorted to the MFA which was a 0.6 mile (12 minute) walk. Upon arrival to the MFA, subjects were also escorted throughout the museum with an MFA work study student.

Subjects walked the perimeter of both floors one and two following a designed route and were asked to stay within those galleries (fig. 1f). Upon entering each gallery, subjects were told to look within the galleries what they found was most interesting. Therefore, each subject might have focused on different artwork. However, there were similarities between subjects in what caught their eye in each gallery. In this version of the experiment, subjects recalibrated every 10 minutes using a recalibration board which was identical to the routine in the main experiment.

Definition of MFA Target and Foil Clips:

The presentation of target and foil clips was identical to that of the main experiment. However, the number of clips seen and how they were annotated slightly differed. Control clips were annotated for using the categories "faces" and "non-faces." Within this variant, subjects were shown 350 foil clips and 350 target clips and like the main experiment (0.05%) of the shots were repeated randomly during the test, for a total of 736 clips shown (excluding one subject who only sat 672 trials). Unlike the main experiment, this variant did not have an equal amount of "faces" and "non-faces" shown. Of the 350 of either target or foil clips about 15% shown were faces and 75% were scenes of the various artwork the subjects were looking at during encoding. All other procedures were identical to the main experiment.

Data Analyses:

The total number of "yes" and "no" responses for each subject were computed using a script in the program MATLAB. However, two subjects were excluded from analyses: one for scoring close to ceiling (96%) due to weather that could not be controlled for and another subject

for having biased responses (responding “yes” <90% of the time) in Version 1. The proportion of yes and no responses was around 50% for each subject included in analyses.

The performance was summarized for each experimental condition where the percentage of trials in which subjects were correct was reported. The overall percentage correct includes the number of target clips the subject responded yes to (true positives) and the foil clips the subject had not seen that they responded no to (false positives) (Tang et al. 2016).

A ROC analysis was done for both versions of the experiment and for the main experiment at both time intervals: 30 hours and 3-4 months. The ROC analysis was a way to look at the false positive rates vs. true positive rates to see the two conditions that constitute a subjects' overall performance.

Shot Content Properties:

To evaluate what factors determine the efficacy of episodic memory formation, a set of content annotations was developed to evaluate each shot and see which aspects of these shots correlate with performance. Subcategories under clips annotated as "faces" included: gender, age, number of faces seen, action, distinctiveness of a person, talking, and if there was interaction with subjects. The subcategories of "non-faces" or "scenes" included: movement, distinctiveness of a scene, or if the scene included people within the background. The performance for each condition was found by summing the total number of true positives and false positives divided by the total number of shots that with corresponding content that was present or not to evaluate performance.

In the MFA variant, the same subcategories for clips with "faces" were used. However, in addition, each category included if a fixation was present along with that condition to see how having a fixation on something particular versus not having a fixation differs in subjects’

performance. For example, in the condition of gender: male vs. female, a further subcategory was used to evaluate the presence of a fixation within that specific segment. Therefore, an example in the case of gender, the annotations were as follows: males with fixation, males without fixation, females with fixations, and females without fixations. These subcategories were used for all of the other conditions as well as gender including age, how many faces were present if the person was executing an action and distinctiveness of a person. The subcategories for "non-faces" or "scenes" included: the distinctiveness of an object, if a painting was present in the shot, or if a statue was present in the shot. The presence or absence of a fixation within these subcategories was also taken into account and performance for all of these conditions was found for each subject and averaged across subjects.

To get the percentages of content shot properties across subjects, each clip a subject saw in a memory recognition test was revisited and further annotated for in both versions. The number of clips the person got correct for a specific condition (ex: gender) was summed and divided by the total number for that specific annotation. For example, if a person saw 65 people who were male in the memory recognition test and got 30 of this clips correct, meaning they remembered seeing that specific person (target) or recognizing they had not seen that person (foil) their score for that particular category would be $(30/65) \sim 46\%$. This was done for each condition and averaged across subjects to get the average performance for a specific content property (fig. 7 and 10).

In the case of the MFA variant, each category (ex: gender) had four conditions instead of two since it took into account if a fixation was present or not. Therefore, for the case of gender (fig. 10a) there is presence of a male (with fixation), male (without fixation), female (with fixation) and female (without fixation). To take fixations into account, there were three considerations: no fixation at all (eye was not able to be tracked), fixation (not on the relevant

item in the clip ex: looking at the floor) and fixation (on the relevant item being annotated for ex: a painting). In the same way as version one to get the average performances the amount of clips that subject got correct for a specific condition (ex: categorization of artwork for paintings) takes the total number of clips there was a fixation on a painting that the subject got correct, over the total number of clips where the subject fixated on a painting in the memory recognition test to constitute performance. These scores were averaged across subjects to get the overall percentage (fig. 10).

Ethical Approval and Informed Consent:

All experimental protocols were approved by the Institutional Board at the Boston Children's Hospital and Massachusetts Institute of Technology. All methods were carried out in accordance with the approved guidelines. Informed consent was obtained from all subjects for all portions and variants of the experiment.

Results

Quantifying Episodic Memories through a Memory Recognition Task: Version 1

The main goal of this experiment was to address long-term episodic memory formation by quantitatively evaluating what humans subjectively filter during a real life encoding task in a memory recognition test that took place at 24 hours up to 4 months post-encoding. The experiment was done in two versions: one indoors (main experiment) and another outdoors (variant 1). In the main experiment, nine subjects walked a route within the Cambridge area including the MIT campus, Kendall Square, and Central Square during a one-hour encoding experience (fig. 1e). Their episodic experience was recorded using a *GoPro* camera and a mobile eye tracker that monitored their eye movements (fig. 1a). Memory for specific content including the two groups of faces vs. non-faces (scenes) was evaluated in two sessions for the main experiment conducted 30 hours and again at 3-4 months post-encoding. The memorability of subjects was evaluated by the presentation of one-second shots of *GoPro* footage from the subjects episodic experience. The shots were defined as “foils” (control clips) and “targets” (from the subjects own experience) which contained annotations of “faces” and “non-faces” (scenes) in equal amounts.

During the recognition memory tasks, target clips were randomly mixed with an equal amount of foil shots, which included shots subjects had not seen and were used as positive controls. The target and foil clips were both controlled for by the following parameters to ensure between subject consistency: same time of day +/- 30 minutes, using the same route and performing the experiment under similar weather conditions to avoid variability and maintain between subject consistency. Subjects then performed a "yes/no" task on the MATLAB program to indicate

whether the subject remembered seeing that particular shot or not and performance was found in number of clips the subjects got correct (fig. 4a).

We expected the performance of subjects on the memory recognition tests to be between chance (50%) and ceiling (100%) somewhere around ~75% based on the previous results found in the Tang et al. 2016 study. However, it was found on average across subjects that performance in the memory recognition test 30 hours post-encoding was 55.7 +/- 3.7 (mean +/-SD) (fig. 4a). This performance was categorized as slightly above chance levels (50%) and well below ceiling (100%). The seven subjects included in this analysis returned 3-4 months post-encoding to complete another memory recognition task. The target clips in both 30 hours and 3-4 months remained the same; however, the foils shown differed from the previous control clips they had seen. It was predicted that when subjects came in 3-4 months post-encoding for the memory recognition task, the average across subjects would decrease since memory decays with time. However, it was found that memory recognition was retained with an average performance at 57.7 +/- 6.8% (mean +/- SD) (fig. 4b). The average performance from 3 hours to 3-4 months had standard deviations that were within the range of each other. To further analyze the data, a t-test assuming equal variances was performed, and it was found that there was no statistically significant difference in the average of the scores ($p < 0.05$).

To see if subjects performed better in one of the main categories of "scenes" vs. "faces" the average number of clips correct labeled as "scenes" and as "faces" were found. It was predicted that subjects were more likely to remember faces, rather than scenes. However, it was found there was no statistically significant difference ($p < 0.05$) in performance of faces or scenes for 30 hours or 3-4 months. On average across subjects the performance of scenes was 55.8 +/- 4.0% (mean +/- SD) and the performance for faces was 55.4 +/- 4.2% (mean +/-SD) (fig. 4c and 4d).

The average performance for scenes on average at 3-4 months post-encoding was 56.4 +/- 5.0 % (mean +/- SD) and for faces 59.2 +/- 8.0% (mean +/- SD). Therefore, there was no statistically significant difference in performance of scenes vs. faces in the memory recognition test 3-4 months post-encoding ($p < 0.05$).

At a further look into the performance of the subcategories of "scenes" vs. "faces" at the two-time points (30 hours and 3-4 months) the performance of faces and scenes per subject was analyzed (fig. 5b and 5c). To further conclude the significance of the data a t-test was performed assuming equal variances, and no statistically significant difference was found in performance for scenes (fig. 5b) or faces (fig. 5c) ($p < 0.05$).

	MAIN	VARIANT 1 (MFA)
Number of subjects:	9(7)	10
Number of subjects tested at 3 months:	7	0
Age range:	18-22	18-22
Age mean \pm SD	20.0 \pm 1.4	20.5 \pm 1.4
Recognition memory tested in one session only	No	Yes

Table 1. Summary of subjects and test conditions in the main experiment and MFA variant. The main experiment and variant 1 (MFA) is described within the methods section. The number of subjects that completed each version of the experiment is indicated, and the number in parentheses is the number of subjects included in data analyses. The overall performance in each experiment is shown in Figure 4a and 4b.

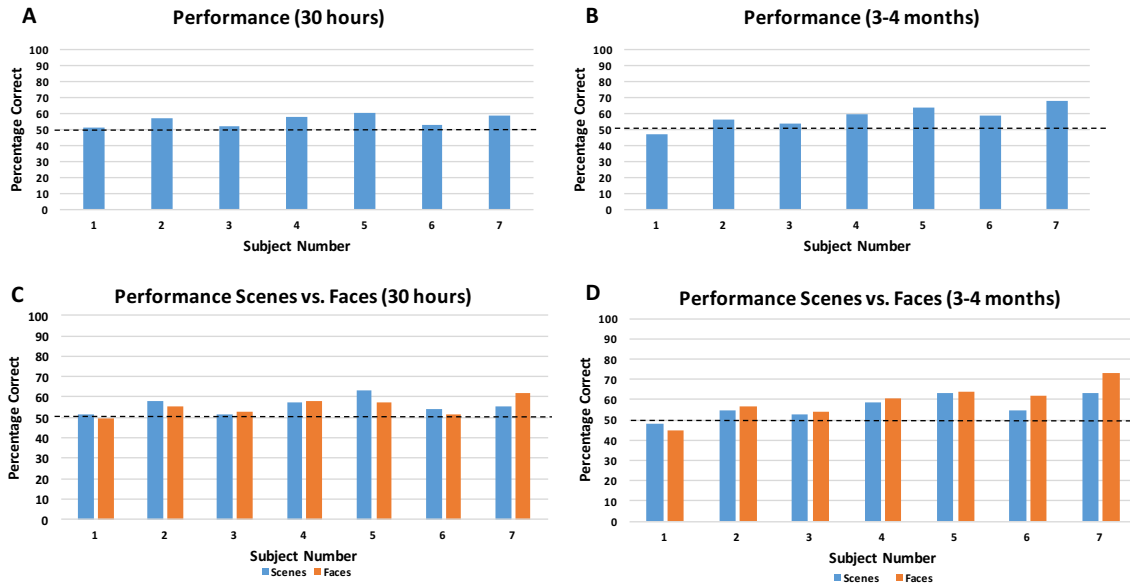


Figure 4. Performance on the memory recognition test was consistent at both time points and roughly equal in the categorization of scenes vs. faces. (A) Overall performance for each subject on the memory recognition test for each subject 30 hours post-encoding. (B) Performance slightly increased on average across subjects 3-4 months post-encoding by 2%, but this increase was not statistically significant ($p < 0.05$). (C) Performance for scenes vs. faces had no statistically significant difference for 30 hours or (D) or for 3-4 months.

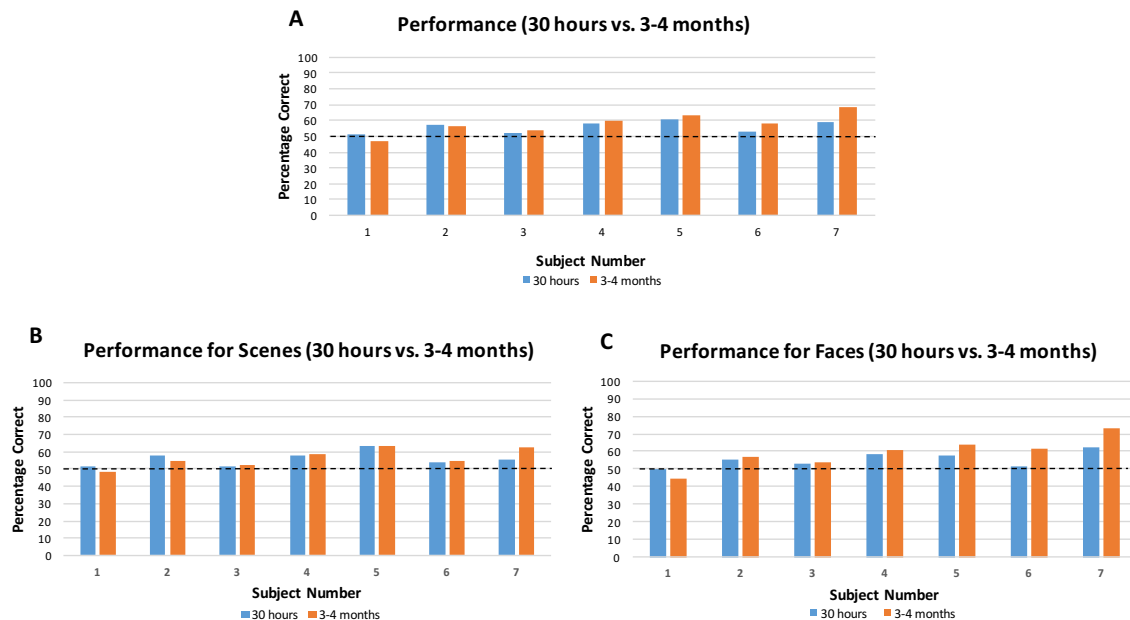


Figure 5. Performance on the memory recognition test was retained across subjects on average 3-4 months post-encoding (A) Overall performance for each subject who completed the memory recognition task 3-4 months post-encoding in comparison with their performance 30 hours post-encoding. (B) Performance between scenes and faces (B) at 30 hours and (C) at 3-4 months remained to have no statistical significance difference in the performance of each group ($p < 0.05$).

To further assess the overall performance in the memory recognition test at 30 hours and 3-4 months, a Receiving Operator Characteristic (ROC) analysis was performed at both time points. This allows for another analysis of the data to see in which of the two conditions considered "correct" subjects are performing better in on average. The performance score on the memory recognition test includes the probability of hits (reporting a correct answer when a target clip was shown) plus the probability of false alarms (reporting an incorrect answer when a foil clip was shown). For the purpose of this experiment, hits are when a subject responds "yes" to a target clip, this is known as a True Positive Rate (TPR), and false alarms are when a subject responds "no" to a foil clip, which is known as a False Positive Rate (FPR). The true positive rate (y-axis) was plotted against the false positive rate (x-axis) on an axis from 0 to 1 in both directions for both 30 hours and 3-4 months (fig. 6a and 6b). These rates were found for each subject at both time points, and each point on the plot represents a subject excluding coordinates (0,0) and (1,1) (fig. 6c and 6d). It was expected that the TPR and FPR values would be roughly equal at (0.5) on average across the seven subjects at both time points. At 30 hours post-encoding, it was found that the average TPR was 0.53 and the FPR was 0.42 or (fig. 6e). At 3-4 months post-encoding the average TPR across subjects was 0.51 and average FPR was 0.35 (fig. 6f). In both scenarios the TPR and FPR values slightly decreased from 30 hours to 3-4 months, therefore in the other two conditions of the ROC analysis false negative rates (FNR) and true negative rates (TNR) increased (fig. 6f). The other two subsequent conditions are known as the false negative rate (FNR), and true negative rate (TNR) include conditions in which the subject was incorrect. False negatives are referring to clips where the subject responds "no" in the memory recognition test, and the clip was from the subjects own experience which was found to be at a rate of 0.47 at 30 hours post-encoding (fig. 6e) and 0.49 at 3-4 months post-encoding (fig.

6f). The other incorrect condition, true negatives, refer to when a subject responds "yes" to a foil. In other terms, thinking they saw a particular person or scene they did not see had a rate of 0.58 at 30 hours post-encoding (fig. 6e) and 0.65 at 3-4 months post-encoding (fig. 6f).

This suggests that the condition which contributes greater to performance on the memory recognition test is when subjects can recognize clips from their experience (TPR) than realizing clips are not from their experience (FPR) at both time points. The analysis also suggests that on average, subjects believe they saw clips that included people or scenes they did not see, proving that episodic memory is malleable.

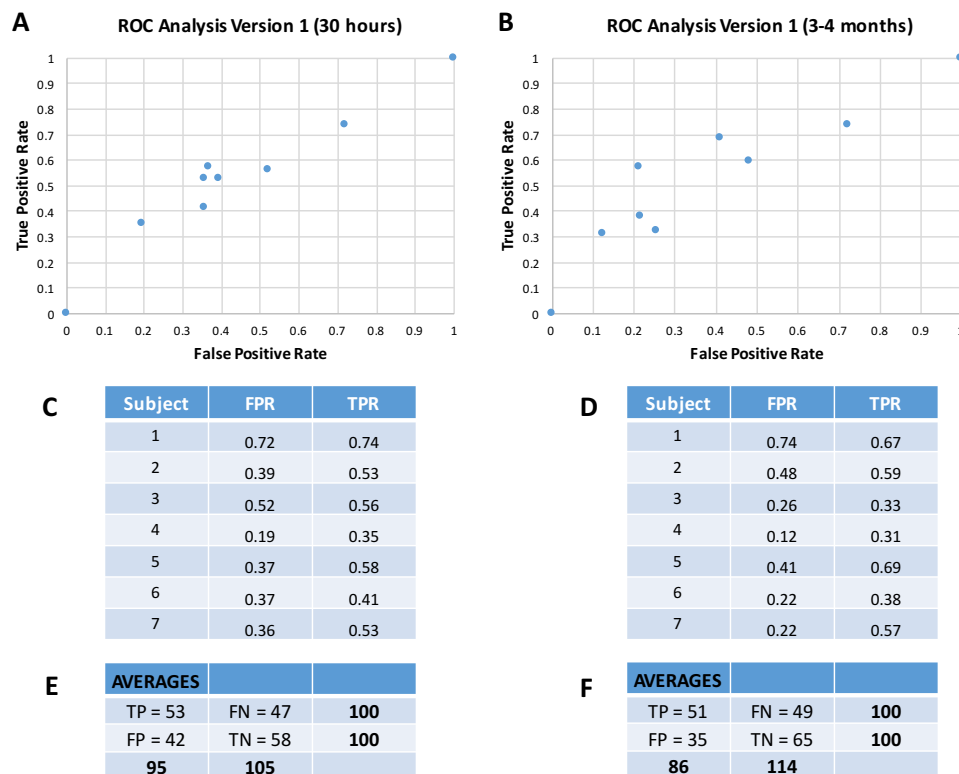


Figure 6. Receiver Operating Characteristic (ROC) Analysis for 30 hours vs. 3-4 months post-encoding showed a decrease in both the False Positive Rate (FPR) and True Positive Rate (TPR). (A) Is a plot of the false positive rate by the true positive rate for 30 hours post-encoding where each dot is a subject except for the threshold dots at (0,0) and (1,1). (B) The same plot but for each subject 3-4 months post-encoding. (C and D) The FPR and TPR for each of the seven subjects included in the overall analysis. (E and F) An average of the FPR and TPR in numbers as it relates to all four conditions at both time points: True Positive (TP), False Negative (FN), False Positive (FP), True Negative (TN).

Since the average scores across subjects on the memory recognition test was slightly above chance (50%) and well below ceiling (100%) we were interested in seeing what content properties of the clips subjects see during their memory recognition test are correlated with higher performance. To do this a set of content annotations was created for the categories of "faces" and "scenes" to assess these properties (refer to methods section). The subcategories in the interest of faces included: gender, age (in relation to subjects, all subjects were between the ages of 18-22), number of faces, presence of action, distinctiveness of a face, if the individual was talking and if the person in the clip was interacting with the subject (fig. 7a-g). The subcategories related to the clips subjects saw labeled as "scenes" included: movement within the scene, distinctive events within the scene (construction scenes, distinctive cars, etc.) and if there was the presence of a person in the background of a scene.

In the case of age and gender, it was predicted that there should be no significance difference in performance whether a male or female was in the clip or if the people were younger or older than the subject. The results in the case of the gender were consistent with the predictions, and it was found there was no statistically significant difference ($p < 0.05$). However, in the case of age (as relation to the subject between the ages of 18-22) it was found that subjects were more likely on average to remember people who were younger than them than older (62.6 +/- 4.5% vs. 55.0 +/- 4.2%) (fig. 7d). In the subcategory of number of faces, it was predicted that if a clip were shown with one face, it would be more likely to be remembered than if multiple faces were shown. However, it was observed that subjects were more likely to remember the clip if multiple faces were present, rather than if only a single face was seen and this difference was statistically significant (58.6 +/- 5.3% vs. 53.7 +/- 4.6%) ($p < 0.05$) (fig. 1c). The subcategory 'presence of action' referred to if the person in the one-second shot was

executing an action such as walking (action) or standing still (no action). It was predicted based on the previous study that shots containing action would be more memorable than shots without action. However, based on these results it was found that when the person in the shot was standing still, rather than executing an action, they were more likely to be remembered during the memory recognition test (61.5 +/- 5.5% vs. 55.0 +/- 4.3%) this difference was considered to be statistically significant ($p < 0.05$) (fig. 1d). In the three remaining subcategories of faces: distinctiveness, if the individual was talking, or if the individual was interacting with the subject during their encoding experience, no statistically significant difference was found ($p < 0.05$). However, the results in each category matched the predictions that a distinctive person, talking, and interacting with the subject would more likely to be remembered than if the person was non-distinctive, not talking or not interacting with the subject (fig. 7e-g). The distinctiveness of a person was considered if there was something about the person that particularly stood out (extremely tall or short, clothing, hair, etc.). In the case of subject interaction, meant if the person in the clip was somehow interacting with the subject by talking to them, or making themselves known to the subject, which would most likely be more memorable to the subject than if a person was simply passing by.

In the subcategories of scenes as mentioned above, it was predicted for movement that movement in the scene would be correlated with higher performance. However, it was found that scenes containing no movement were remembered slightly higher on average (fig. 7h), but this difference was not statistically significant ($p < 0.05$). It was predicted that if a distinctive scene would be correlated with higher performance on the memory recognition test and the results matched this prediction (fig. 7i) with distinctive scenes having an average performance across subjects of 58.7 +/- 4.8% and non-distinctive scenes at 52.6 +/- 4.2%. This result was proven to

be statistically significant ($p < 0.05$). As for the final subcategory of scenes assessed: presence of people in the background, it was predicted that if people were in the background of a particular scene, this would be a cue that would enhance memory recognition. Although this result was found (fig. 7j) there was no statistically significant difference ($p < 0.05$).

These results suggest that we filter out a lot that happens in our daily lives, whether it be people we see or specific scenes from an hour of our day. The four categories of statistical significance found within the shot content properties correlated with performance on average across the seven subjects provides insight for what was more memorable during the experience of these subjects within this particular setting outdoors in Cambridge.

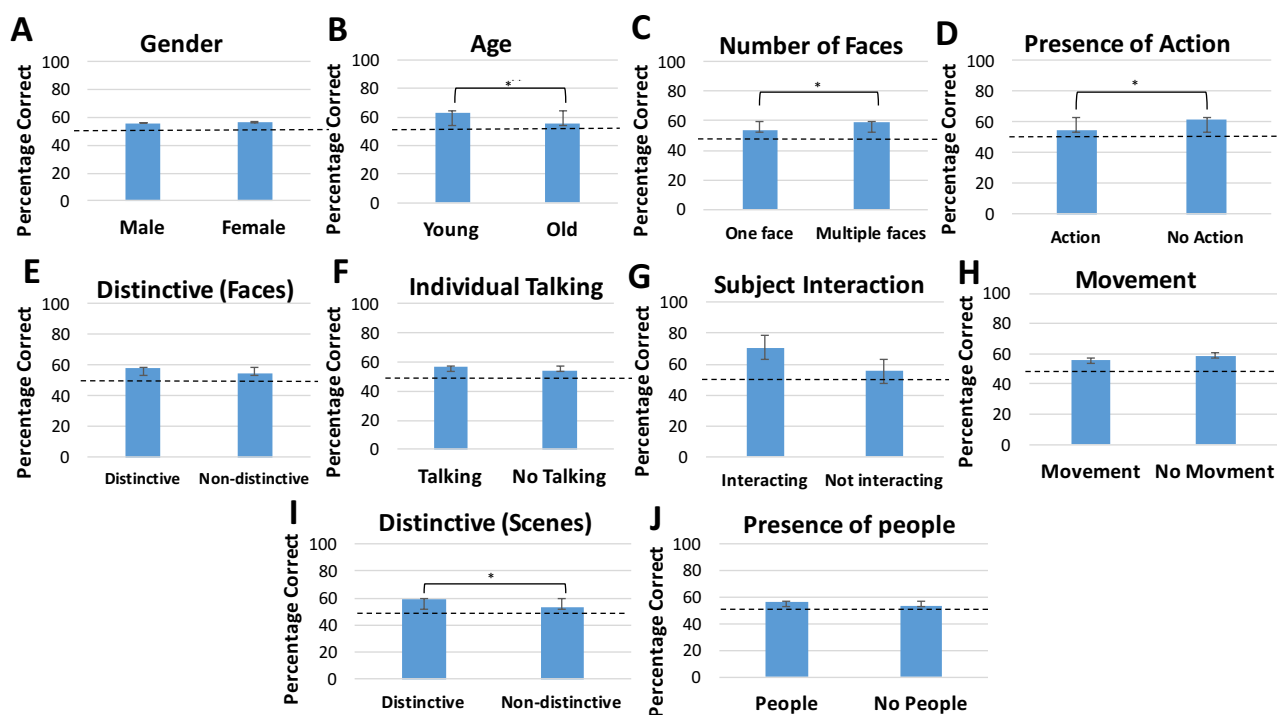


Figure 7. Average Shot Content Properties correlated with performance across the subject pool (A-J). Performance for the memory recognition test that corresponds to if the content was present in the shot or not (mean +/- SEM). A quantitative analysis was performed within the major categories of "faces" and "scenes" to measure shot content properties with performance. (A-G) are subcategories of "faces." (H-J) are subcategories under "scenes." Of various content shot property categories, four were found to be statistically significant ($p < 0.05$) including: age of the person in the shot (A), number of faces present in the shot (C), if the person in the shot was executing an action (D) and if there was a distinctive or non-distinctive scene (I).

Quantifying Episodic Memories through a Memory Recognition Task: Version 2

The results from the first version of this experiment offered a promising start to studying real-world episodic memory formation and retrieval. To see memory formation in a different context, and how fixations on certain objects or people enhance memory recognition a second version was performed indoors at the Museum of Fine Arts (MFA) with ten subjects. The same procedures were followed as in version one of the experiment, however, since the settings were different the number of faces subjects saw was significantly less than the outdoor route (refer to the methods section). Therefore, the categories of "faces" and "scenes" subjects saw during the memory recognition test were not in equal amounts and consisted mostly of scenes within the museum. The number of "target" and "foil" clips remained equal in the memory recognition test, and the same parameters were controlled for to maintain between subject consistency (time of day +/- 60 minutes and the museum route).

Based on the results of version one of this study it was predicted that the average performance of the subjects would fall somewhere between 55% (average across subjects of version one) and 75% (midway between chance (50%) and ceiling (100%)). It was found that the average performance across subjects was 62.3 +/- 7.0% (mean +/- SD) for the ten subjects included in analysis (fig. 8a). This performance was within the expected range based on previous data results.

To assess the performance in the categories of scenes versus faces the percentage correct for each was calculated and reported. Based on the results of version one of this experiment, it was predicted that there would be no significant difference in the performance scenes versus faces. The results found that on average subjects were correct in clips shown with faces 65.3 +/- 12.0% (mean +/- SD) of the time and correct in clips with scenes 61.7 +/- 7.0% (mean +/-SD) of

the time (fig. 8b). There was no statistically significant difference in performance of scenes versus faces ($p < 0.05$).

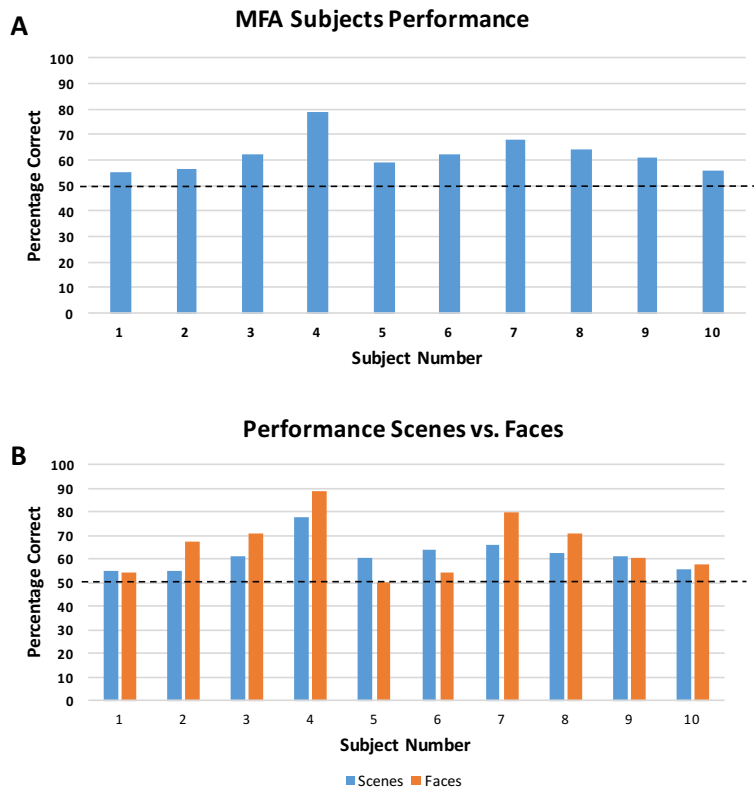


Figure 8. Performance in the MFA variant was consistent with version one and had no statistically significant difference in scenes vs. faces (A) Overall performance for each subject on the memory recognition test for each subject 24 hours post-encoding with an average of 62.3 +/- 7.0 (mean +/-SD). (B) Performance for scenes vs. faces had no statistically significant difference ($p < 0.05$). Faces: 65.3 +/- 12.0 Scenes: 61.7 +/- 7.0.

To further assess the overall performance in the memory recognition test for version two, a Receiving Operator Characteristic (ROC) analysis was also performed as it was in version one. This allows for another analysis of the data to see in which of the two conditions considered "correct" subjects are performing better in on average. The true positive rate (y-axis) was plotted against the false positive rate (x-axis) on an axis from 0 to 1 in both directions. These rates were found for each subject at both time points, and each point on the plot represents a subject

excluding (0,0) and (1,1) (fig. 9b). It was expected that the TPR and FPR values would be roughly equal at (0.5) on average across the ten subjects, which was also predicted in version one. It was found that on average that the true positive rate across ten subjects was 0.61 and the false positive rate was 0.37 (fig. 9c). The other two subsequent conditions are known as the false negative rate (FNR), and true negative rate (TNR) include conditions in which the subject was incorrect. False negatives are referring to clips where the subject responds "no" in the memory recognition test, and the clip was from the subjects own experience which was found to be at a rate of 0.39 (fig. 9c). The other incorrect condition, true negatives, refer to when a subject responds "yes" to a foil. In other terms, thinking they saw a particular person or scene they did not see had a rate of 0.63.

This suggests that the condition which contributes greater to performance on the memory recognition test is when subjects can recognize clips from their experience (TPR) than realizing clips are not from their experience (FPR). The analysis also suggests that on average, subjects believe they saw clips that included people or scenes they did not see, supporting the idea that episodic memory is malleable.

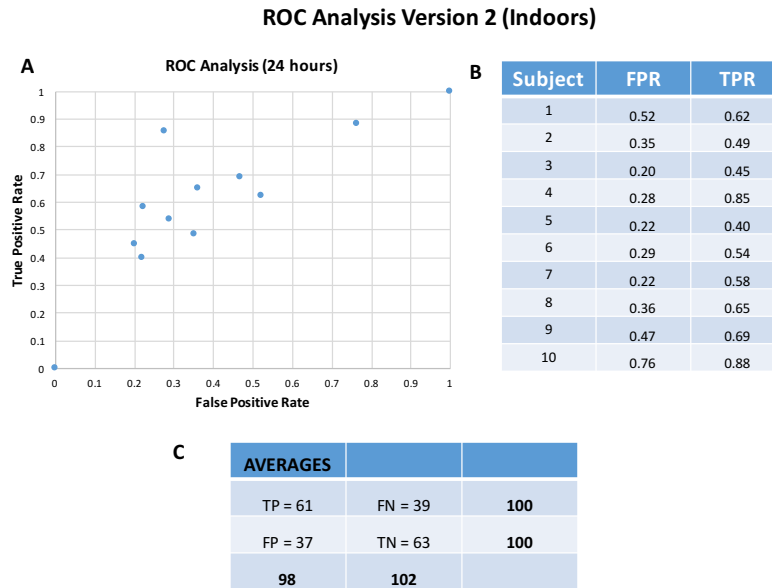


Figure 9. Receiver Operating Characteristic (ROC) Analysis for the MFA variant. (A) Is a plot of the false positive rate by the true positive rate for 30 hours post-encoding where each dot is a subject except for the threshold dots at (0,0) and (1,1). (B) The FPR and TPR for each of the ten subjects included in the overall analysis. (C) An average of the FPR and TPR in numbers as it relates to all four conditions: True Positive (TP), False Negative (FN), False Positive (FP), True Negative (TN).

To further assess the content properties correlated with higher performance on the memory recognition test as in version one, a similar set of content annotations was created for the categories of "faces" and "scenes" to assess these properties. Within each category, the relevance of a fixation as detected by the eye-tracking software on the particular person, object, or artwork was also taken into account to see how fixations can enhance memory recognition. It was predicted that if a person fixated on something or someone shown in a clip, they would be more likely to remember it on the memory recognition test and have a higher performance on these specific clips. It was found that on average across the ten subjects that this was the case in most categories of faces or scenes. All categories looked at for each condition took into account if a fixation was present or absent in that particular clip (fig. 10). The categories looked into for

faces were as follows: gender (fig. 10a), number of faces present (fig. 10b), age (in relation to subject between the ages of 18-22) (fig. 10c), if the person was executing an action or not (fig. 10d), if the person was talking (fig. 10e), and distinctiveness of an individual (fig. 10f). The following subcategories for scenes were as follows: distinctiveness of an object (fig. 10g) and the categorization of artwork (whether it was a statue or a painting) (fig. 10f).

To further assess the significance of these results a one-factor, ANOVA was performed for each condition to measure the variance between the four groups of each condition (example: gender) ($p < 0.05$). It was found that after performing an ANOVA for each group the subcategories of faces: talking (fig. 10e) and for scenes: if an object was distinctive or not (see fig. 10g). Since the ANOVA only tells you if the means of groups differ, a t-test assuming equal variance was performed to further assess the significance of the data ($p < 0.05$). It was found that the categories of statistical significance included if an individual being fixated on was talking (fig. 10c) and if a distinctive object was fixated on (fig. 10g). In the case of fixations on individuals, it was found that subjects on average performed better on the memory recognition test if they fixated on a person who was talking during their encoding experience (90.7 +/- 0.19%) than if they had not fixated on them (73.0 +/- 0.10%). Similarly, a person was more likely to remember seeing a distinctive object within the museum if they fixated on that particular object (76.8 +/- 0.13%) versus if they had not fixated on the distinctive object (59.4 +/- 0.09%).

These results suggest that on average subjects were more likely to remember a particular face or scene when they fixated upon something during encoding, which was later shown in the memory recognition test.

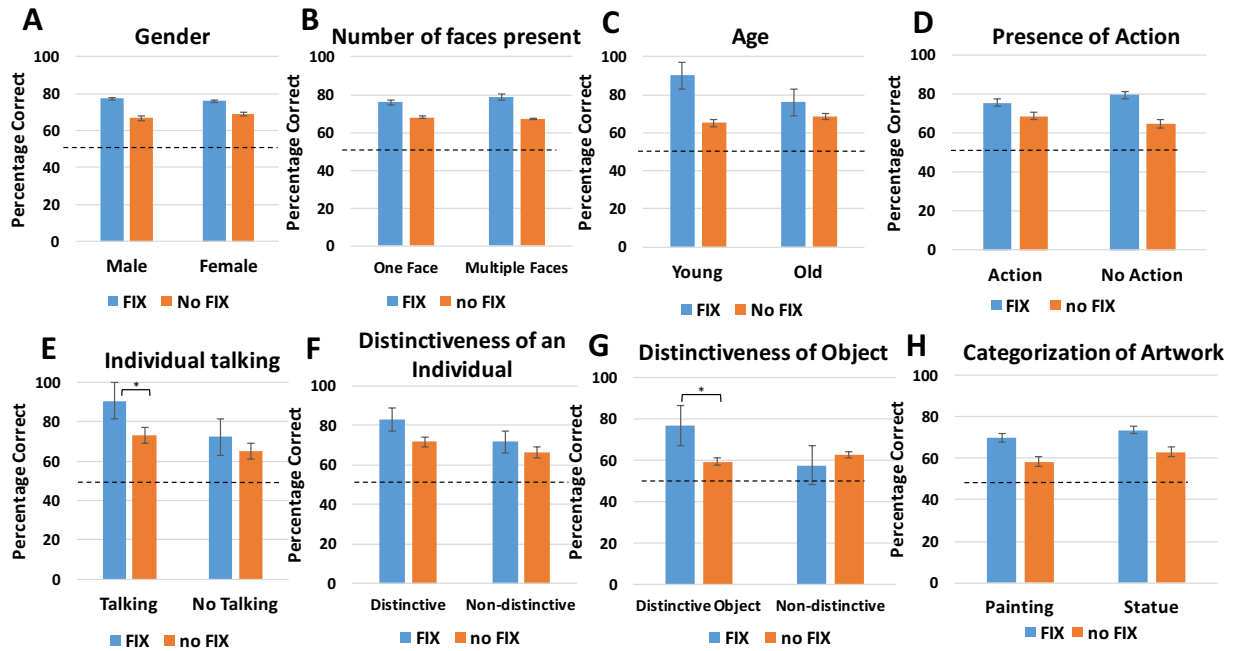


Figure 10. Shot Content Properties correlated with performance (A-H). Performance for the memory recognition test that corresponds to if the content was present in the shot or not (mean +/- SEM). A quantitative analysis was performed within the major categories of "faces" and "scenes" to measure shot content properties with performance. The presence of a fixation was taken into account and categorized in the memory recognition performance along with shots that did not include fixations (A-F) are subcategories about "faces." (G and H) are subcategories under "scenes." Of the various content shot properties if an individual the subject fixated on was talking (E) vs. if they were talking and the subject had not fixated on them, had a statistically significant difference ($p < 0.05$). In the subcategory of scenes, if the person fixated on a distinctive object (G) vs. if they had not fixated on that object, had a statistically significant difference ($p < 0.05$).

Discussion

Episodic memories constitute a minuscule fraction of what people remember in the real world. They include isolated events that took place in someone's life and are full of detail, yet extremely malleable. It was previously hypothesized that during one's episodic experience specific contents relevant for episodic memory formation would include notable events that occurred or specific faces they encountered to determine the efficacy of human memory. Within this study, it was found that we are exposed to a lot of sensory stimuli impinging upon our brains in our everyday lives, and we make an account to remember so very little of it. This was seen in both versions of the study: indoors and outdoors, and included the both encoding sessions from the main experiment (30 hours and 3-4 months) outdoors. Regarding looking into what aspects were more memorable to subjects, the content properties which were annotated for in both the outdoor and indoor version allowed for a deeper insight into the efficacy of long-term memorability. The categories which showed statistically significant differences ($p < 0.05$) in performance matched our predictions in version one where distinctive scenes were more memorable with a higher average performance than ones considered less distinctive (fig. 7i). In the case of the indoor version where the presence of fixations was taken into account, it was predicted that fixations would be correlated with a higher performance on the memory recognition test. This prediction was supported in the analyses when it was seen that on average in most categories our subjects were more likely to remember particular scenes or faces if they had fixated on a particular person or object that was later shown in the memory recognition test (fig. 10).

The findings of this study extend previous knowledge within the field because most studies have focused mainly on in lab memory studies where lists of pictures or isolated words

were given as constructs of episodic memories (Bahrick 1975, Rubin 1996, Brady 2008, Vogt 2007, Standing 1973, Castelhana 2010, & Andermane 2015). These studies have addressed these questions and identified cues to the efficacy of human memorability. However, these attempts at evaluating episodic memories lack realistic context. To study episodic memory formation under natural conditions, it is vital to study the temporal and spatial context that are critical to real memories. To effectively tackle the understanding of real-life episodic memory formation, the Kreiman lab had previously taken an interesting alternative to examining memory formation using movies as stimuli (Tang et al. 2016). The overall findings of this study had supported previous work in the field and showed that of the 161 subjects that participated, accuracy was as high as 80% in memory recognition tasks (Tang et al. 2016). Within my study, it had been found that subjects were only able to recall on average up to 62.3 +/- 7.0% of their episodic experience on average which opens up new findings that had not been seen previously.

There are several explanations as to why real-life memories are less memorable than lists of words or pictures (Bahrick 1975, Rubin 1996, Brady 2008, Vogt 2007, Standing 1973, Castelhana 2010, & Andermane 2015) or the episodic memory formation in movies (Tang et al. 2016). It is important to note that although movies are close to real-life stimuli, they are not real life. Movies are meant to be memorable regarding what actors are saying, how they dress and how they act. The study was a promising start to quantitatively examining episodic memory formation in individuals, however, the artificial stimuli within a movie are manipulated by the writers and directors (Tang et al. 2016). Movies make every attempt to manipulate a person's attention and feelings, which ultimately effects their recollections of these experiences (Tang et al. 2016). Therefore, the conclusions from this study regarding the predictability of episodic memory formation were not able to be extrapolated to real life episodic memory where movies

are not used as stimuli. Another important aspect to note is that within the Tang et al. study, like all other studies on episodic memory in the field, they took place in a laboratory. Therefore, the movie watching was isolated within a room where other things would not going on besides movie watching. In real life, a lot of things are happening at once, especially in the outdoor route around Cambridge. In this way, within the movie version of this study, your attention is solely focused on the one stimuli, whereas in real-life there is too much context to filter through and remember every detail. This selective filtering process and constructive processes formed by subjects are important to note and have been seen in previous behavioral neuroscience contexts (Loftus 2005).

Because subjects were scored on a numeric performance scale and the performance was constructed of two conditions of being "correct" it is also important to see the conditions in which subjects are not correct and a possible explanation for why this is. In the conditions where subjects are "incorrect," they can be thinking they saw people or scenes they did not see or not remembering aspects of their experience they did see. In the condition of subjects believing they saw people or scenes, they did not say, we are referring to the false negative rate (FNR). Explanations for this and biological basis for why this occurs can be explained by the idea of misinformation implanted in a person's brain after the encoding process, which is the first stage of the memory process (Loftus 2005). In the case of subjects not remembering people or scenes they did see, we refer to this as the true negative rate (TNR) which would be known as "forgetting." Previous literature has tried to explain these aspects of why people forget the details within episodic memories. Recent literature has tried to consider an implication for the behavioral and neurobiological models of episodic memory retention and forgetting. It is thought that loss of specific details that occur during the encoding portion of memory formation can

reflect both storage and retrieval deficits (Sekeres 2016). Therefore, disruptions in any stages of the human memory process: encoding, storage or retrieval can lead to the loss or forgetting of specific details within these memories.

It has been seen for the past century that there are multiple mechanisms researchers have tried to identify as to why people forget details of memories. Some of these theories include memory decay (Thorndike 1913) and interference during learning (Keppel and Underwood 1962). Because episodic memories, in particular, are autobiographical events, they contain a lot of context information including sensory, perceptual, and affective information (Tulving 1972). Therefore, they are specifically susceptible to the loss of detail information (Tulving 1972). Previous studies of the retention of episodic memories in healthy controls found that the details encoded and retention of these details decline with time (Furman et al. 2007, 2012). In those with hippocampal damage, such as patient H.M. previously discussed have deficits in forming new episodic memories. This has highlighted the role of the hippocampus in the different memory systems we possess as humans (Rosenbaum 2005).

With these ideas of forgetting, it is important to address the factors and details we do remember to categorize what is more likely to be forgotten. Researchers have pointed out that that previous studies done in the field with lists of words or pictures are not clear if these findings would apply to more complex, naturalistic autobiographical memories (Sekeres 2016). My study tried to address the naturalistic stimuli we perceive as humans and see what details within these episodic memories are retained. In this way, the aspects that subjects do not perform as well on can be considered as details we are more likely to forget.

Upon a quantitative analysis of the specific details, subjects remembered it was found that specific aspects were more statistically significant in the performance of subjects on the

memory recognition test. When subjects were shown specific faces the contents of age, number of faces present and if the person was executing an action were categories of statistical significance ($p < 0.05$). In the case of age, subjects were more likely to remember a person if they were considered younger than the subject performing the experiment. One explanation for this is because walking around the city in Cambridge or at a museum you are less likely to see a younger child by themselves, which can be a more distinctive cue for the individual. In the case of number of faces, subjects performed better when there were multiple faces within a shot versus if there was only one face present. This could be because there are more aspects a subject can remember if multiple people were present or a large group versus just seeing one person. In the context of action, subjects were more likely to remember the face of the person if they were standing still and not executing an action. This could be explained by the fact that there is a greater chance of a subject encoding a specific face if they were standing still versus if they were moving and the subject might not be paying attention since they are just simply passing by. In the context of scenes, subjects performed better when a distinctive scene was shown within their route in Cambridge. The idea for this can be explained by the fact that a more distinctive scene like a fire truck passing by or a brightly colored car that would cue the attention of an individual is more likely to be encoded and retrieved than a nondistinctive scene. Data from a recent study supports these findings because a distinctive element of a memory would be considered a central element, and is less likely to be as susceptible to decline within episodic memory retrieval (Sekeres 2016).

Also, the results found in version two with relation to eye fixations on particular objects, faces, or scenes, we had seen a higher performance on average across subjects in most categories. This finding supports previous literature which found that memory performance on

immediate and long-term memory tests was dependent on a fixation of a critical object at the time memory encoding took place (Castelhano & Henderson 2010). The categories found to be statistically significant ($p < 0.05$) for performance on the memory recognition test included distinctive objects the subjects had fixated on to be greater than if the subject had not fixated on that particular object. In the content property category of faces which had statistical significance ($p < 0.05$) in performance on the memory recognition test, it was found subjects were more likely to remember a person they had fixated on that was talking, rather than if the person was talking and the subject had not fixated on them. These results suggest that as humans we are more drawn to looking at particular people or objects that are distinctive and draw our attention. When we fixate on these people or objects, we are more likely to remember them, than if we had not had a fixation on that particular person or object.

Based on our conclusions of our study, we were able to see that in-lab studies are not directly translated to those in real-world experiences. Since studies testing memory recognition in the field had not previously been done outside of a laboratory setting to our knowledge, we were able to see that although episodic memory formation might have been strong in past studies, when you take the situation into a real-life context that is not the case. It is important to note at the same time previous literature and studies within the field of neuropsychology provide a basis for understanding why people mistake their experiences with subjective constructs within their minds (Loftus 2005). This phenomenon is referred to as the "misinformation effect" which has been identified by memory researchers in the past (Okado and Stark 2005). It is primarily associated with the impairment people have in memory and how this leads to us believing that we had seen something or someone we had not (Loftus 2005). In regards to my study, subjects were scored on their performance in a memory recognition test and at a further examination of

results through a ROC analysis the rates at which subjects were considered incorrect was determined through the true negative rates (TNR) and false negative rates (FNR). In the condition of a true negative rate, subjects are answering "no" to people or scenes they did see and in the case of false negative answering "yes" to people or scenes they did not see. In all variants of the memory recognition experiment, the true negative rates were between 58 and 62%, and the false negative rates were between 38 and 48%. Therefore, it can be seen that as humans we have a distortion as to what and who we remember supporting previous literature (Loftus 2005).

Some implications for as to why this occurs has been studied and seen the biological mechanism as to why this can exist. It is thought that this misinformation can become incorporated into a memory already encoded and stored within the hippocampus (Greene et al. 1982). In the second process of memory formation within the hippocampus, storage, the human brain filters, selects and retains the encoded data from the neural code (Straube 2012). This aids in the explanation as to why a lot of what happens during or daily lives is forgotten. Additionally, this can lead to an altered memory because this misinformation becomes incorporated in the previously encoded memory. Specific memory traces can be susceptible to decay with time and alteration of these memories (Loftus 2005).

This also relates back to the ideas of learning and memory previously discussed. Because memory cannot occur without learning any disruption in the learning process can lead to the memory being misconstrued (Loftus 2005). Because we are constantly being bombarded with external stimuli in our daily lives, other cues in the environment can lead to a person not encoding another specific aspect of what they experience. Whether this is a person they passed,

construction scene, fire truck or anything that a person might experience while walking through a city or in a museum.

It is thought that there is a specific nature to misinformation of events and that some people can be more subjected to false memories than others. It has been found that young children and the elderly are more likely to susceptible to misinformation than adults (Ceci and Bruck 1993). The subjects within my study ranged from the ages of 18-22, and are therefore considered young adults and fall somewhere within the range of children to adults. This could provide an explanation as to why the true negative and false negative rates are high, in the case that subjects thought they saw things they did not see. Regardless of this, it is important to note that at any age misconstruction of what did happen can occur. However, there are certain age groups that are more likely to be affected than others.

This study has been an interesting basis to study behavioral neuroscience in a real-life context and not within an artificial laboratory setting. It is important to note that as there are flaws in other studies of episodic memory within laboratory settings, there are also several flaws associated with running an experiment outside of a controlled setting. Because our study took place outside of a laboratory: outdoors (version one) and indoors (version two), there are a lot of things that cannot be controlled for. For example in both versions of the experiment, how many people the subjects pass during their encoding experience or the presence of something interesting or distinctive that could enhance memorability across subjects. Although timing and routes were controlled for, what happens within the period of encoding cannot be controlled for. Therefore, it can be difficult to ensure that each subject has the same experiences to test memorability on. Additionally, the context of the outdoor route and the indoor route in the museum were vastly different as to what the subjects were exposed to. In the outdoor route,

subjects were exposed to lots of people since it took place in busy parts of Cambridge but lacked a lot of interesting, distinctive scenes to test subjects memorability on. In version two at the MFA, there was much fewer people present at the time subjects completed encoding and were tested mostly on pieces of artwork or distinctive objects they had seen in the museum. As a next step, it would be interesting to complete this study in a setting that would ensure seeing a roughly equal amount of people and interesting scenes.

Another implication would be to run a larger sample size because although the results were consistent across subjects, there was a small sample size of 10 subjects, which might not be translated to a larger group that completed the same study. Additional future directions for this study could include investigating the neuronal underpinnings of episodic memory formation (Kreiman 2007) and expanding the study to longer time scales. Furthermore, it would be beneficial to investigate other forms of evaluation such as free recall and task dependency. Lastly, another future implication for this study would understand how emotional aspects contribute to real world memory formation in humans.

Conclusion

This study has expanded upon previous knowledge in the field by providing a way to study episodic memory formation under natural conditions. Previous studies within the field have focused largely on in-lab recollections of words, faces, objects, or scenes. These studies lack spatiotemporal context that is necessary for evaluating episodic memory formation under real-world conditions. This study has allowed us to quantify memorability of real-life episodic experience by using ground truth data from each subject and subsequently evaluating their ability to recognize those events.

The results of this study found real-life episodic memories to be less memorable than the previous study in the lab using movies as stimuli (Tang et al. 2016) with an average performance across subjects up to 62.3 +/- 7.0%. When studying episodic memory formation outside of a laboratory, it can be seen that at any given moment a multitude of stimuli is impinging upon our senses. Only a small fraction of these inputs crystallize to form episodic memories. Extensive work within the field of episodic memory has proven that the memories we store only are made up of a small number of input signals. The human brain constructs a narrative based on sensory inputs that it selects and interprets.

This study has allowed us to quantitatively evaluate the aspects that have enhanced long-term memory by identifying properties that enhance memorability in the real world. To the best of our knowledge, this work constitutes the first quantitative approach to directly measure memory formation with ground truth data in real life scenarios.

Acknowledgements

I would like to thank Dr. Gabriel Kreiman for allowing me the opportunity to complete my senior thesis on episodic memory within his laboratory and helping me write my final thesis.

I would also like to thank Matt Peterson of the Kanwisher lab at MIT for allowing me to use his experimental design of recording real world visual experience and for all of his help in the computer programming end and borrowing the equipment to complete both versions of my study.

I would also like to thank Lynn Courtney in the head of education at the museum of fine arts and the help of her work study students who helped us make the second version of this experiment possible by completing it at the MFA.

References

1. Tang, H, Singer, J, Ison, M, Pivazy, G, Romaine, M, Frias, R, Meller, E, Boulin, A, Carroll, J.D., Perron, V, Dowcett, S, Arlellano, M, Kreiman, G (2016). Predicting episodic memory formation for movie events. *Scientific Reports*, 6, 30175.
2. Squire, L. R., Stark, C. E. & Clark, R. E (2004). The Medial Temporal Lobe. *Annual Review of Neuroscience* 27, 279-306.
3. Peterson, M.F., Lin J., Zaun I., Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of Vision*. 16(7):12,1-18.
4. Blumberg, J., and Kreiman, G. (2010). How cortical neurons help us see: visual recognition in the human brain. *Journal of Clinical Investigation* 120:3054-3063.
5. Tulving, E. (2002). Episodic memory: from mind to brain. *Annual review of psychology*. 53, 1-25
6. Schacter, D.L., Norman, K.A. & Koutstaal, W (1998). The cognitive neuroscience of constructive memory. *Annual review of psychology* 49, 289-318.
7. Schacter, D.L. & Slotnick, S.D. (2004). The cognitive neuroscience of memory distortion. *Neuron* 44, 149-160.
8. Loftus, E.F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learning and Memory* 12, 361-366.
9. Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A. & Rosenbaum, R.S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current opinion in neurobiology* 16, 179-190.
10. Schmidt, S.R.(1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of experimental psychology. Learning, memory, and cognition* 11, 565-578.
11. Gabrieli, J.D.E. (1998) Cognitive Neuroscience of Human Memory. *Annual Review of Psychology*. 49:87-115.
12. Kreiman, Gabriel. (2007). Single unit approaches to human vision and memory. *Current Opinion in Neurobiology*. 17:417-475.
13. Castelano, M., Henderson, J. (201). Incidental Visual Memory for Objects in Scenes. *Visual Cognition*. 12:6, 1017-1040.
14. Andermane, N., Bowers, J.S. (2015). Detailed and gist-like visual memories are forgotten at similar rates over the course of the week. *Psychon Bull Review*. 22:1358-1363.
15. Huang, Y., & Mucke, L. (2012). Alzheimer Mechanisms and Therapeutic Strategies. *Cell*, 148(6), 1204–1222
16. Smarr, B. L., Jennings, K. J., Driscoll, J. R., & Kriegsfeld, L. J. (2014). A Time to Remember: The Role of Circadian Clocks in Learning and Memory. *Behavioral Neuroscience*, 128(3), 283–303.
17. Cohen, G., & Conway, M. A. (2008). Memory in the real world. Hove: Psychology Press.
18. Bahrick, H. P., Bahrick, P. O. & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of experimental psychology: General* 104, 54.
19. Straube, B. (2012). An overview of neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions*. 8:35.

20. Rubin, D. & Wenzel, A. (1996). One hundred years of forgetting: a quantitative description of retention. *Psychological Review* 103, 734–760.
21. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Science USA* **105**, 14325–14329 (2008).
22. Vogt, S. & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology* 54, 298–303.
23. Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology* 25, 207–222.
24. Loftus, E. F. & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist* 35, 409–420.
25. Greene, E., Flynn, M.S., and Loftus, E.F. (1982). Inducing resistance to misleading information. *J. Verbal Learn. Verbal Behav.* (21): 207–219.
26. Ceci, S.J. and Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. *Psychol. Bull.* (113): 403–439.
27. Okado, Y. and Stark, C.E.L. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learn. Mem.* (12): 3–11
28. Suzuki, W. A. (2007). Making new memories: the role of the hippocampus in new associative learning. *Annals of the New York Academy of Sciences* 1097, 1–11.
29. Sekeres, M. (2016). Recovering and preventing loss of detailed memory: differential rates of forgetting for detail types in episodic memory. *Learning & Memory* **23**:72–82.
30. Thorndike EL. (1913). *Educational Psychology* (Vol. 2). Teacher's College, Columbia University, New York.
31. Keppel G, Underwood BJ. (1962). Proactive inhibition in short-term retention of single items. *J Verb Learn Verb Behav* (1): 153–161.
32. Furman O, Dorfman N, Hasson U, Davachi L, Dudai Y. (2007). They saw a movie: long-term memory for an extended audiovisual narrative. *Learn Mem* (14): 457–467.
33. Furman O, Mendelsohn A, Dudai Y. (2012). The episodic engram transformed: time reduces retrieval-related brain activity but correlates it with memory accuracy. *Learn Mem* 19: 575–587.
34. Rosenbaum RS, Köhler S, Schacter DL, Moscovitch M, Westmacott R, Black SE, Gao F, Tulving E. (2005). The case of K.C.: contributions of a memory-impaired person to memory theory. *Neuropsychologia* (43):989–1021.
35. Eichenbaum, H. (1997). How does the brain organize memories? *Science* **277**, 330–332.
36. Estes, W.K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Reviews* 104, 148-169.

