Emotion modulates eye movement patterns and subsequent memory for the gist and details of movie scenes

Ramanathan Subramanian
Advanced Digital Sciences Center (ADSC)
University of Illinois at Urbana-Champaign, Singapore

Divya Shankar
Centre for Mind/Brain Sciences and Department of Cognitive Sciences
University of Trento, Italy

Nicu Sebe
Department of Computer Science and Information Engineering (DISI)
University of Trento, Italy

David Melcher
Centre for Mind/Brain Sciences and Department of Cognitive Sciences
University of Trento, Italy

A basic question in vision research is where people look in complex scenes and how this influences their performance in various tasks. Previous studies with static images have demonstrated a close link between where people look and what they remember. Here, we examined the pattern of eye movements when participants watched neutral and emotional clips from Hollywood-style movies. Participants answered multiple-choice memory questions concerning visual and auditory scene details immediately upon viewing one-minute long neutral or emotional movie clips. Fixations were more narrowly focused for emotional clips and immediate memory for object details was worse compared to matched neutral scenes, implying preferential attention to emotional events. While we found the expected correlation between where people looked and what they remembered for neutral clips, this relationship broke down for emotional clips. When participants were subsequently presented with key-frames (static images) extracted from the movie clips such that presentation duration of the target objects corresponding to the multiple-choice questions was matched, and the earlier questions were repeated, more fixations were observed on the target objects and memory performance also improved significantly, confirming that emotion modulates the relationship between gaze position and memory performance. Finally, in a long-term memory test, old/new recognition performance was significantly better for emotional scenes as compared to neutral scenes. Overall, these results are consistent with the hypothesis that emotional content draws eye fixations and strengthens memory for the

doi: Received: December 30, 2013

scene gist while weakening encoding of peripheral scene details.

Keywords: emotional movie clips, eye movements, short-term memory, long-term memory, attention

Introduction

A classic finding, going back to seminal work by Buswell (1935) and Yarbus (1967) is that the pattern of eye movements when viewing a complex scene is driven by both bottom-up and top-down factors. When looking at complex scenes such as photographs or pictures, participants typically look around the image through an initial exploratory stage with many short fixations that are largely driven by bottom-up salience, followed by a later stage with a slower pace of fewer saccades per second and an increasing influence of the task (Buswell, 1935; Yarbus, 1967; Tatler, Hayhoe, Land, & Ballard, 2011).

More recently, these eye movement patterns have been linked to performance on immediate and long-term memory tests (Melcher & Kowler, 2001; Hollingworth, Williams, & Henderson, 2001; Tatler, Gilchrist, & Land, 2005; Pertzov, Avidan, & Zohary, 2009). First, it is generally reported that participants are better at remembering items that were fixated (Melcher & Kowler, 2001; Hollingworth, 2006). Second, studies of memory for pictures and photographic images have provided evidence that observers accumulate information about the visual details of scenes over time and across separate glances (Tatler & Melcher, 2007; Pertzov et al., 2009; Nuthmann & Henderson, 2010; Tatler et al., 2011). However, the level of detail which can be reported about a scene depends on the stimulus complexity and task (Tatler & Melcher, 2007; Tatler et al., 2011). In addition, emotional content in the stimulus can modulate memory for scene details (Melcher, 2010), perhaps by influencing where participants look in the image (Calvo & Lang, 2004; Kaspar et al., 2013).

Static images of scenes differ from real-world experience in many ways (see Tatler et al., 2011 for a review) raising the question of how the dynamic nature of natural viewing might influence where we look and what we remember. One way to bridge the gap between photographs and the real world is to study movies. Recently, there has been a renewed interest in the use of movies as stimuli which provide complex visual stimulation within a naturalistic situation (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Hasson et al., 2008; Hasson, Malach, & Heeger, 2010; Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Soleymani, Pantic, & Pun, 2012). Recent studies suggest that the pattern of fixations is particularly consistent across observers for Hollywood-style movies (Tosi, Mecacci, & Pasquali, 1997; Goldstein, Woods, & Peli, 2007; Dorr, Vig, & Barth, 2012; Wang, Freeman, Merriam, Hasson, & Heeger, 2012; Smith & Mital, 2013). This finding suggests that when watching movies, observers are strongly guided by the narrative and editing techniques used by professional film-makers. If attention and gaze are driven by these cues, then fixation patterns may differ greatly from those typically studied with static scenes and free viewing.

Given the important link between where people look and what they remember, studies of eye movements with movies raise the question of how gaze position and memory are linked when watching movies. Studies of visual memory have measured performance for home-made movies, which offer more control over the stimulus, and with Hollywood-style movies. One approach has been to measure change-detection by altering the identity, position or visual properties of an object during a *cut* (the end of a continuous shot from a single camera). For example, (Hirose, Tatler, & Regan, 2010) measured change-detection for color, position, identity or shape changes. They found that when the object property was changed across a cut, memory was biased towards information presented after the cut. In another study, subjects were shown a sitcom during an fMRI experiment and then, after a delay ranging from a few hours to 9 months, they were given a surprise memory test (Furman, Dorfman, Hasson, Davachi, & Dudai, 2007; Hasson et al., 2008). Specifically, memory for the narrative content of the movie was tested using recognition and cued recall. In addition to questions about the plot of the film (such as social interactions or jokes) there were questions about object details, such as "What type of sandwich did Larry offer the homeless man?". Performance on such questions was initially quite high after three hours but then decreased dramatically over time.

In contrast, memory for the plot or social relationships stayed relatively robust across the 9-month period (Furman et al., 2007).

One important aspect of Hollywood-style movies is that they are designed to entertain viewers in order to be popular and to sell tickets. One of the main ways to achieve this goal is to evoke an emotional response from the audience. Indeed, certain movie genres, such as *adventure*, *horror*, *romance* or *comedy*, are expressly defined by the types of emotions they elicit in viewers. This makes movies an interesting and valuable stimulus, since within each movie there will be clips which are visually quite similar but are designed to evoke different emotions. In a typical romantic comedy, for example, there are some scenes which are relatively neutral in emotion, but also ones that the director hopes will evoke laughter or tears. Based on previous studies with static images, it is possible to make some predictions about how emotional content in movies might influence eye movements. In the short term, emotional images are thought to draw attention, eye movements and more processing resources compared to neutral ones (Calvo & Lang, 2004; Attar, Andersen, & Müller, 2010). Thus, one might expect more narrowly focused gaze positions when comparing across observers viewing an emotional movie clip.

In terms of what people remember after viewing the movie clip, numerous studies suggest that emotion boosts memory for the gist of the event rather than the details (see Buchanan & Adolphs, 2002 for review). A recent study with emotional photographs (Melcher, 2010) found a difference between emotional and neutral pictures, with memory for object details being worst for negative emotion images. However, the emotional item itself (*e.g.*, snake or gun) may be remembered better (Kensinger, Garoff-Eaton, & Schacter, 2007) as compared to peripheral, non-emotional items. It has also been suggested that while recollection of contextual details for emotional scenes may not be very accurate, emotion nevertheless enhances the subjective feeling of remembering (Rimmele, Davachi, Petrov, Dougal, & Phelps, 2011).

The aim of the current experiment was to investigate how emotion influences where people look when watching movies, and link this to subsequent memory for the viewed content. Participants viewed movie clips that were approximately one minute long, and were then presented with a number of written questions regarding visual details of the clips as well as the spoken narrative. Fixated positions were then used to compare memory for viewed objects in an immediate memory test for scene details. We hypothesized that emotional content would focus the viewers strongly on those objects responsible for the emotion, causing them to ignore peripheral visual details. While all movies, to some extent, can lead to poor memory for non-central details (Hirose et al., 2010), we expected strong emotions to exacerbate this trend. A correlation between fixated scene regions and memory for details regarding corresponding scene objects (or target objects) was observed for neutral movie clips but not for emotional clips, confirming our hypothesis.

In order to directly compare our findings to previous studies on eye movements for static images, eight weeks after the original experiment, key frames (still images) from the movie clips were shown to participants for the same duration as the target objects corresponding to the posed questions had been visible, and questions identical from the original experiment were repeated. Significant differences were observed between the eye movement patterns for movie clips and static images, in the form of shorter fixations, longer saccades and greater dispersion of fixations across the scene area for static images, consistent with other recent studies (Dorr et al., 2010; Smith & Mital, 2013). Also, considerably higher number of fixations were observed on the target objects on which the memory questions were based, and participants were significantly better at recalling details of static images. Experimental results comparing eye movements and memory performance of participants for movie clips and static images (where the emotion perceived in the movie stimulus was absent) confirm that emotion modulates the relationship between where people look and what they remember. At the same time, we additionally tested subjects' 'old/new' recognition for clips in order to study the role of emotion in long-term memory for scene gist. While the focus on emotional narrative might prove costly in terms of short-term recollection of scene details, we expected that it would instead strengthen the gist memory trace (Kensinger et al., 2007) and make such clips seem more familiar (Rimmele et al., 2011) in a long-term memory test. Consistent with this expectation, old/new recognition performance was significantly better for emotional clips as compared to neutral clips. Cumulatively, the observed results suggest that emotion limits the gazing and encoding of scene details, while strengthening memory for scene gist.

General Materials and Methods

Participants

There were 24 students aged between 19 and 40 ($\mu = 24.9$) with normal or corrected-to-normal vision who participated in the study. Observers provided written informed consent, as approved by the Ethics committee on involving human subjects at the University of Trento. The participants were paid either a small fees or given course credits for participation.

Materials

Ten Hollywood-style movies were used for the experiment: Airplane (Jim Abrahams, David & Jerry Zucker, 1980), August Rush (Kirsten Sheridan, 2007), The Gods Must Be Crazy II (Jamie Uys, 1990), Legally Blonde (Robert Luketic, 2001), Life is Beautiful (Roberto Benigni, 1997), Love Actually (Richard Curtis, 2003), Remember the Titans (Boaz Yakin, 2000), Slumdog Millionaire (Danny Boyle, 2008), The Truman Show (Peter Weir, 1998) and Up (Bob Peterson & Pete Docter, 2009). A majority of these films were chosen based on a previous study which aimed at creating a dataset of emotional movie clips (Bartolini, 2011). Three movie clips- one each of *neutral*, *negative* and *positive* valence were chosen from every movie. These movie clips were used as stimuli, and were roughly of 1 minute duration ($\mu = 64.7s; \sigma = 22s$). Prior to running the experiment, each of the clips was rated by 3 referees for their emotional value (most suitable emotion tag and valence). The valence rating provided by the referees for each movie clip closely matched that given by participants during the study- Figure 1 presents a scatter plot of the participants' valence-arousal (VA) ratings, confirming that the movie clips did elicit the expected emotions and that subjects had little difficulty in determining the stimulus valence. A description of the movie clips used in the experiment and the most suitable emotion tag for the movie clip, as identified by a majority of the participants, is provided in Table 1.

Table 1: Movie clips shown in the experiment. *Emotion tag* denotes the most suitable emotion type for a movie clip, as determined by the majority of participants.

S.No	Movie	Duration (min:sec)	Description	Emotion Tag
1	**Airplane**	0:48	Air-hostess Elaine leaves her love, ex-pilot Ted.	*Sad*
2		0:38	Elaine introduces a boy to the flight captain and co-pilot in the cockpit.	*Neutral*
3		1:25	Reactions of a worried woman and her co-passengers as Ted struggles to control the aircraft.	*Funny*
4	**August Rush**	1:26	A mother enquires about her lost son at the counselor's office.	*Sad*
5		0:57	The lost son- a musical prodigy, talks to a vagrant musician.	*Neutral*
6		1:30	The son meets his parents while performing at a concert.	*Happy*

7	**Gods must be crazy II**	0:58	Two African tribal siblings are separated as one falls off from a water tank trailer in which they travel.	*Sad*
8		0:29	A pilot learns from a colleague about an approaching desert storm even as his friend and a woman take off in a chopper.	*Neutral*
9		1:05	The couple return back safely in the chopper after some embarrassing moments and a desert adventure.	*Funny*
10	**Legally blonde**	0:51	Elle- an aspiring lawyer, gains admission to Harvard Law School.	*Happy*
11		0:37	Elle converses with her classmate, Vivian.	*Neutral*
12		1:16	Elle is shocked when her professor makes sexual advances on her.	*Sad*
13	**Life is beautiful**	0:57	Funny and charismatic Guido arrives at Dora's (his love) school posing as an education officer.	*Funny*
14		0:34	Dora inadvertently gets into Guido's car and starts talking to him.	*Neutral*
15		1:54	Guido gets shot by a Nazi soldier even as he makes his son believe that they are playing a game.	*Sad*
16	**Love actually**	0:51	Introductory scene purporting that 'Love actually is everywhere'.	*Happy*
17		1:02	Juliet knocks on Mark's studio to ask if he can play her wedding ceremony videotape.	*Neutral*
18		0:58	Juliet realizes from the video that Mark is in love with her.	*Sad*
19	**Remember the Titans**	1:00	New coach speaks to the *Titans* football team on fighting hatred at a cemetery marking the Battle of Gettysburg.	*Neutral*
20		1:19	One of the key *Titans* players is paralyzed by a car accident.	*Sad*
21		0:52	*Titans* win the football championship.	*Happy*
22	**Slumdog Millionaire**	0:44	Conversation between Latika and Jamal at her workplace.	*Neutral*
23		1:09	Latika is kidnapped by gangsters when she comes to meet Jamal at the railway station.	*Sad*
24		1:20	Latika and Jamal reunite at the railway station.	*Happy*

25		1:29	Truman's love is driven away from the beach by her father.	*Sad*
26	**The Truman Show**	0:33	Truman tells his friend that something strange is going on around him.	*Neutral*
27		1:36	Truman leaves the show and the audience react excitedly.	*Happy*
28		1:00	Carl- a shy quiet boy meets the energetic Ellie, an adventure enthusiast.	*Funny*
29	**Up**	0:50	Ellie tells Carl about Paradise Falls and Charles Muntz, the famous explorer.	*Neutral*
30		1:29	Ellie (now old) falls ill and dies.	*Sad*

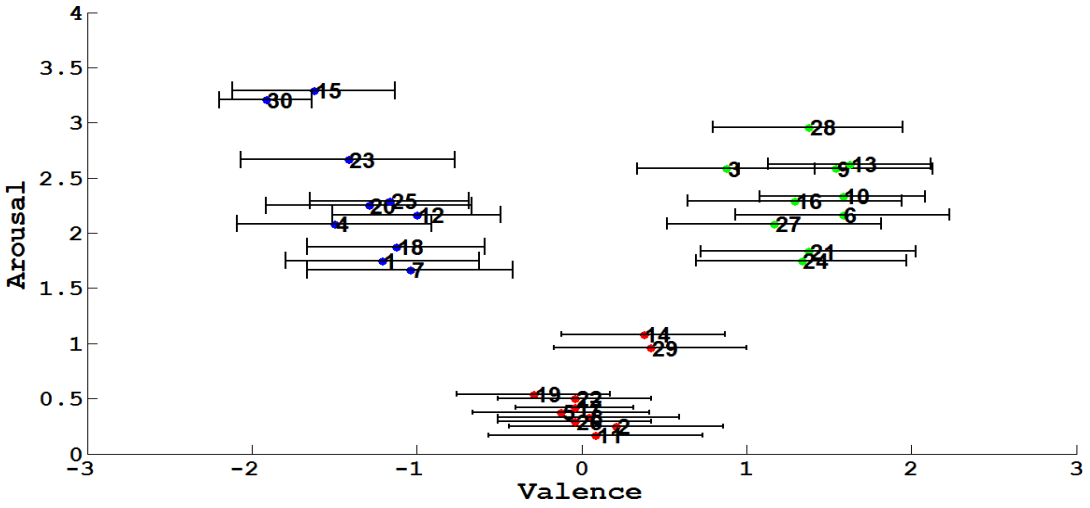


Figure 1: Scatter plot showing participants' mean valence and arousal ratings for the 30 movie clips. *Neutral*, *positive* and *negative* valence stimuli are plotted using red, green and blue colors respectively. Individual valence ratings range from -2 to 2, while arousal ratings are from 0–4. Errorbars denote one standard deviation along the valence direction. Clip IDs are as in Table 1.

The movie clips dimensions ranged from 616×256 to 704×352 pixels, and were presented to participants on a 17" LCD monitor placed approximately 60 cm from their seated position thereby subtending a maximum visual angle of 35° in both horizontal and vertical directions. While participants watched movie clips, their eye movements were tracked using the Tobii T120 desktop eye tracker. The tracker records eye positions with respect to a 1280×1024 pixel resolution screen every 8.3 ms (120 Hz sampling frequency), and is accurate to within 0.4° visual angle upon nine-point calibration under these conditions. The minimum fixation duration and saccade thresholds were set to 100 milliseconds (ms) and 6 pixels/ms (50 pixels/sample) during the recordings. Finally, since faces are known to be visually attractive, we checked if there were any differences in the number of faces per frame and their size on screen for the different emotion conditions using the popular Violoa-Jones face detection software (Viola & Jones, 2004). We computed the number of faces and their sizes (as a fraction of the display area) for the different emotion conditions. The mean number of faces/frame was found to be $1.68 \pm 0.52, 1.7 \pm 0.99$ and 0.86 ± 0.55 and the average size of faces was $0.07 \pm 0.04, 0.09 \pm 0.07$ and 0.07 ± 0.05 for neutral, positive

and negative movie clips respectively. The face sizes were not significantly different for the various emotions as confirmed by t-tests ($p > 0.05$ in all cases).

Experiment 1- Immediate Memory (IM) Task

Procedure

Participants watched each of the 30 movie clips on a computer screen while sitting in a dimly lit room. After each clip, instructions appeared on the screen asking them to (i) determine the most appropriate emotional tag for each movie clip, (ii) rate the clip for valence and arousal and (iii) answer the multiple-choice questions pertaining to visual or auditory details of the viewed clip. The most suitable emotional tag was to be chosen from a set of six options- *angry*, *funny*, *sad*, *happy*, *neutral* or *none of the others*. The valence scale ranged from -2 (most negative) to 2 (most positive), while arousal options ranged from 0 (remained indifferent or calm) to 4 (reacted intensely to the stimulus). Upon viewing each movie clip, participants were presented with multiple choice questions pertaining to the visual and conversational details in the clip to test their memory (see Figure 2). These questions regarded peripheral scene objects not central to the narrative. For example. questions regarded the ***color*** ("what color was the dress worn by the woman with the gun: *black*, *white* or *blue*?"), ***identity*** ("what are the contents of the truck: *wooden sticks*, *animal tusks* or *tree trunks*?") or ***location*** ("where is the fire extinguisher: to the *left of the door*, to the *right of the door* or on the *shelf next to the man*?") of certain objects in the movie clip stimuli. Some questions also concerned the ***content of spoken conversation*** ("what does the woman want to become: a *doctor*, *lawyer* or *engineer*?"). Since most movie clips involved conversations between scene characters and we hypothesized that faces would attract significant visual attention, none of the IM questions concerned faces and facial characteristics. A total of 60 questions were presented- 20 each relating to details from neutral, positive and negative clips. 13 questions concerned spoken dialogue, while the remaining focused on visual details. These questions were presented on the computer screen in large text and participants responded using a mouse button to select the answer they deemed as appropriate. Participants were instructed to choose one response, with three possible options presented for 55 of 60 questions and two options for the remaining five questions. The experiment lasted approximately one hour. An overview of the experimental procedure is presented in Figure 2.

Data Analysis

One of the challenges in using real movies as stimuli is that each clip differs from others in numerous ways. Based on previous literature, we identified seven factors of each clip which might influence memory performance.

1. Movie clip length (L_c): Although longer viewing time might give participants more time for memory encoding, it also increases the amount of information which might need to be remembered and put more emphasis on long-term memory rather than working memory. Thus, we hypothesized that longer movie clips might lead to worse performance.

2. Number of shots in the movie clip (NS_c): as calculated using the standard shot-detection software `http://shotdetect.nonutc.fr/`. Given prior studies (*e.g.*, Hirose et al., 2010) showing poor change detection across cuts, we predicted that a clip made up of multiple shots (and cuts) would lead to worse memory for object details.

3. The visibility duration of target items (VD_t) in the movie clip (expressed as a proportion of the entire clip length): Target objects on which the IM questions were based were visible for a median of 6.8 seconds, and the visibility duration ranged from 626 ms to 60 s. We expected that when the target object was visible over a longer period of time, participants were more likely to correctly answer questions about that object.

4. Time at which the target objects disappear from the scene (TD_t) (expressed as a proportion over the entire clip length): In a given movie clip, the target object(s) may be shown in the beginning or towards the end of the clip. This could lead to *primacy* or *recency* effects, such that objects shown either at the beginning or end of the clip are remembered better.

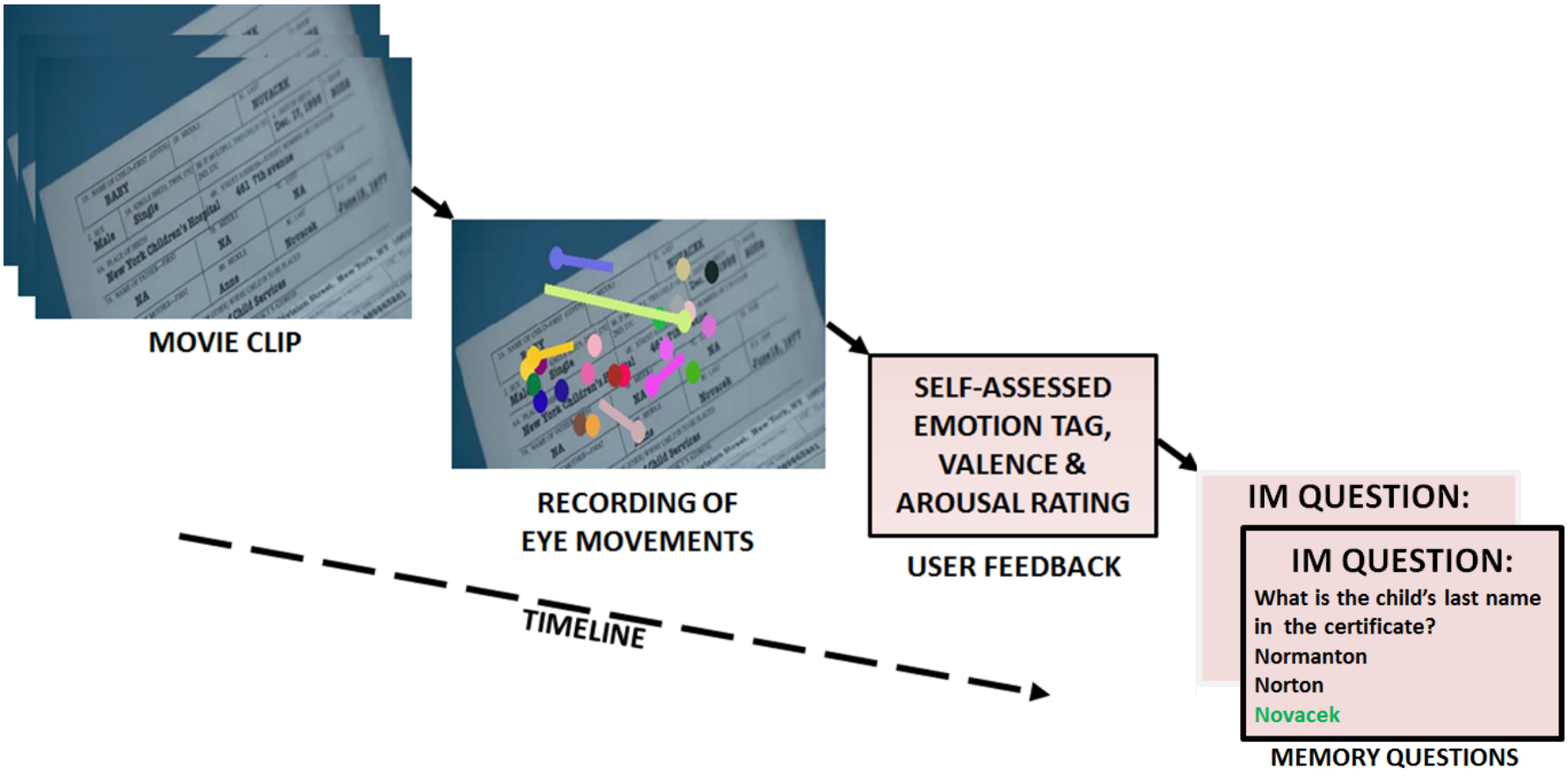


Figure 2: Illustration of the experiment protocol: In each trial, participants viewed a movie clip of roughly 1 minute duration. Their eye movements were recorded even as they viewed the stimulus. Upon viewing, participants were required to determine the most appropriate emotional tag, as well as rate the clip for emotional valence and arousal. Also, they were asked a series of Immediate Memory (IM) questions relating to scene details- *e.g.*, identity, color, location of specific scene objects or pertaining to the conversational details in the clip. Frames from the movie 'August Rush' are used for illustration with permission by Warner Bros. Entertainment Inc. (All Rights Reserved).

5. Length of emotional content within the movie clip (LE_c): the beginning and end of the most emotional portion of each movie clip was annotated by three independent observers. Typically, the onset of the most emotional part of each clip was preceded by some neutral background content in order to effectively evoke the target emotion. It seemed possible that if the duration of the most emotional part of the clip was short, then emotion would have minimal impact on the the memory for object details.

6. Clip valence (V_c) as rated by the participants: whether the movie clip was rated as *positive*, *neutral* or *negative* was expected to influence memory for scene details and conversation, as described above (see Introduction). Particularly, in the case of negative emotion clips, we expected that attention would be focused on the central (emotional) aspects of the scene rather than object details.

7. Clip arousal (A_c) ratings provided by participants: we predicted that increased arousal would lead to more focus on the gist of the movie and thus poorer memory for object details.

Analysis of Target Objects

To directly analyze the relationship between fixated positions and performance in the memory test, we marked rectangles around the target objects (TOs) on which the IM questions were based in each video frame, over the duration in which they were visible. On average, these rectangles respectively occupied $8.2 \pm 7.2\%$, $7.6 \pm 6\%$ and $7.8 \pm 8\%$ of the frame area for the neutral, positive and negative stimuli. As the location of TOs with respect to the scene center is known to impact fixation likelihood, average TO distance from the center was matched across the different valence stimuli. On average, centers of the TO rectangles were found to be 0.484, 0.454 and 0.401 semi-diagonal length times away from the frame center for neutral, positive and negative movie clips respectively. Post-hoc t-tests confirmed that the TO sizes and distances from the scene center for the different emotion conditions were not significantly different.

A user was assumed to fixate on a TO in a particular frame if the eye fixation fell within the corresponding rectangle– the proportion of frames in which the TOs were fixated on over their visible duration was $7.8 \pm 18\%$, $6.5 \pm 14.4\%$ and $6.6 \pm 11.8\%$ respectively for neutral, positive and negative valence stimuli. As expected, the proportion of eye-fixations on each TO as compared to other TOs correlated positively (Pearson's r) with the TO rectangle size for neutral ($r = 0.97, p < 0.0000001$), positive ($r = 0.93, p < 0.0000001$) and negative stimuli ($r = 0.94, p < 0.000001$).

Results

We tested whether emotional content influenced the pattern of eye movements by calculating (i) mean fixations/second, (ii) mean fixation duration and (iii) mean saccade amplitude for each subject over the different emotional clips. Consistent with our hypothesis, there were interesting differences in eye movement patterns for the different emotion types. The highest number of fixations/second were observed while viewing positive valence movies (1.65), followed by neutral (1.51) and negative stimuli (1.45). Paired post-hoc t-tests showed that the fixation rate differences were significant between neutral and positive ($t_{23} = -2.2136, p < 0.05$), as well as positive and negative ($t_{23} = -3.531, p < 0.005$) stimuli. However, the difference between neutral and negative stimuli was not significant ($t_{23} = -1.0066$, n.s.). Conversely, mean fixation duration was longest for neutral scenes (546.5 ms), followed by negative (541 ms) and positive (496.6 ms). Post-hoc paired t-tests again showed that the differences were significant between positive and negative stimuli ($t_{23} = 4.8686, p < 0.0001$), and neutral and positive ($t_{23} = 3.94, p < 0.001$), but not between neutral and negative ($t_{23} = -1.00$, n.s.).

Saccade amplitudes across stimuli were computed as a fraction over the frame diagonal length (so that the largest saccade amplitude is 1). A larger mean saccade amplitude was observed for neutral stimuli (0.1637) as compared to positive (0.1539) and negative (0.1502) movie clips. A one-way ANOVA test confirmed the significant main effect of emotion on saccade amplitudes ($F_{(2,71)} = 4.19, p = 0.0192$). Post-hoc paired t-tests showed significant saccade length differences between neutral and positive stimuli ($t_{23} = -4.3313, p < 0.0005$), as well as neutral, negative stimuli ($t_{23} = -5.6903, p < 0.00001$). The saccade lengths for positive and negative movie clips however, did not differ significantly $t_{23} == -1.214$, n.s.).

To verify if emotional clips induced more focused gaze positions among viewers, we employed the *entropy* measure proposed in (Judd, Ehinger, Durand, & Torralba, 2009). Entropy provides an estimate of (i) the scene breadth covered by eye fixations of a particular viewer, and (ii) the consistency with which the viewer population fixates at particular scene locations. Since the IM task required participants to sample and analyze scene details, one would normally expect subjects to attempt to cover as much as the scene as possible. However, emotional scene objects in the scene could limit this tendency to wander around the scene, resulting in a lower spread of eye fixations over the scene by each user, and consequently, greater coherence in the scene locations fixated by the population.

To study if emotional stimuli modulated gaze behavior, we aggregated eye fixations over each *video shot*, which consists of a series of contiguously captured pictures and therefore, constitutes the atomic representation of a movie scene. Then, we convolved a Gaussian filter over the fixated locations to synthesize the continuous shot saliency map (CSSM) and computed the shot entropy (SE) as $-\sum_X(p(x)log_2(p(x))$, where $\{X\}$ denotes the set of gray values in the CSSM, and $p(x)$ denotes the probability distribution of each $x \in X$. Upon resizing all CSSMs to 200×100 pixel resolution, we compared the mean entropy over shots (MES) computed from (i) eye fixation aggregates for each viewer, and (ii) eye fixation aggregates for all viewers, for the various emotional stimuli.

Figure 3 presents the CSSM generated from the eye fixations of a particular viewer and the viewer population for one neutral and emotional stimulus respectively. As expected, our analysis confirmed that the MES was higher for neutral stimuli, both within and across subjects. Concerning how a particular viewer fixated on the various emotional stimuli, the per-viewer MES was higher for neutral clips (MES =2.83), as compared to positive (MES =2.02) and negative (MES =2.27) valence clips. A one-way ANOVA test confirmed the main effect of emotion on the MES score ($F_{(2,71)} = 47.62, p < 0.000001$). This finding was further reinforced by post-hoc paired t-tests, showing a significant difference in the MES score between neutral and positive ($t_{23} = 17.9033, p < 0.000001$), neutral and negative ($t_{23} = 12.1413, p < 0.000001$) as well as positive and negative ($t_{23} = -6.532, p < 0.000005$) stimuli. Considering the eye fixation characteristics of the entire set of participants, maximum dispersion of fixated locations over shots was again observed for neutral stimuli (MES = 6.06), followed by negative (MES = 5.14) and positive stimuli (MES = 4.86). Here, paired t-tests revealed a significant difference between neutral and positive stimuli ($t_9 = -3.4327, p < 0.01$), marginally significant for neutral vs negative ($t_9 = -2.2442, p = 0.0515$), and an insignificant difference between positive and negative stimuli ($t_9 = -0.8340$, n.s.).

A crucial difference between using entropy for analyzing spread of eye fixations over images and video shots is that the latter can include artifacts in eye movement patterns due to motion of scene objects/camera. In this regard, the study by (Dorr et al., 2010) in which eye movements were compared for Hollywood movie trailers, hand-captured natural videos and static images observed that eye fixations for movie trailers were the most centered and coherent, in spite of involving the maximum motion. Therefore, camera motion in movies is manipulated to focus viewers' attention on specific scene objects, usually appearing around the screen center. Since we based the IM test questions on peripheral scene details, viewers were forced to disperse their visual attention away from the central narrative– we hypothesized that such dispersion was more possible for neutral scenes. To confirm this hypothesis, we compared the average inter-frame motion (denoted using pixels/frame) for the different emotional stimuli. The mean inter-frame motion for the neutral, positive and negative movie clips were found to be 17.4, 35.3 and 31.4 respectively, and left-tailed post-hoc t-tests revealed that the motion in neutral stimuli was significantly less compared to positive ($t_9 = -2.2889, p < 0.05$) or negative ($t_9 = -1.8601, p < 0.05$) stimuli. Even though neutral stimuli involved the least motion and were generally shorter than emotional clips (Table 1), highest MES was observed for neutral clips. This observation suggests that the observed difference in the spread of fixations was due to emotional valence rather than low-level factors such as object/camera motion artifacts.

Next, we related the pattern of eye movements to the ability of participants to remember the details of the movie clips. As seen in Figure 4, performance was significantly better in the neutral condition (57.3% mean accuracy over all question types) compared to the two emotional conditions (43.8% for positive and 45% for negative valence stimuli). This was consistent with our hypothesis that the tendency to focus only on central details relevant to the scene narrative in movies would be intensified by emotional content. A two-way repeated measure ANOVA with 'emotion category' and 'question type' as factors revealed a main effect of 'question type' ($F_{(3,59)} = 10.29, p < 0.0001$), while the main effect of 'emotion category' was marginally significant ($F_{(2,59)} = 3.1, p = 0.0542$).

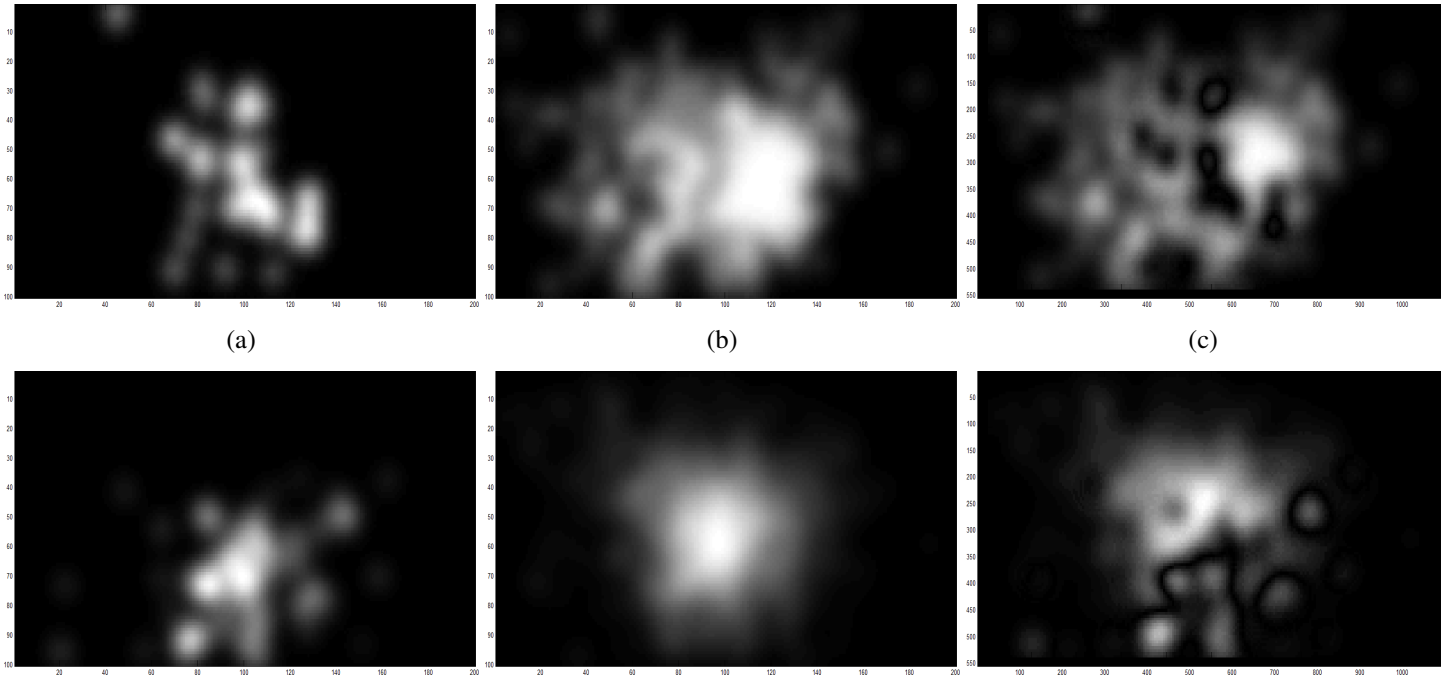


Figure 3: (Top row) Mean CSSM for Participant 14 (MES = 3.04), and all viewers (MES = 6.68) for the neutral clip from 'Gods Must be Crazy II' comprising two shots. Differences between the two maps are shown in (c). (Bottom row) Mean CSSM for (a) Participant 14 (MES = 2.75), (b) viewer population (MES = 5.61) and (c) difference map for the sad clip from the same movie composed of 37 shots. All maps are resized to 200×100 pixels.

Also, the interaction between 'question type' and 'emotion category' was not significant ($F_{(6,59)} = 1.13$, n.s.). Post-hoc t-tests revealed a significant difference in memory performance between neutral and positive clips ($t_{38} = -2.06, p < 0.05$), while the difference between neutral and negative clips was somewhat significant ($t_{38} = -1.83, p = 0.075$). Similar memory performance was observed for positive and negative clips ($t_{38} = -0.1945$, n.s.).

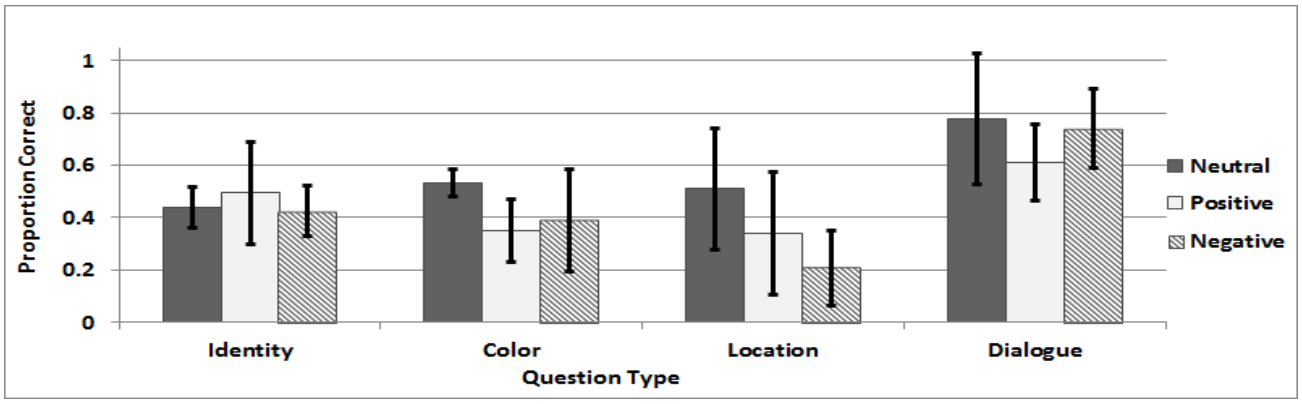


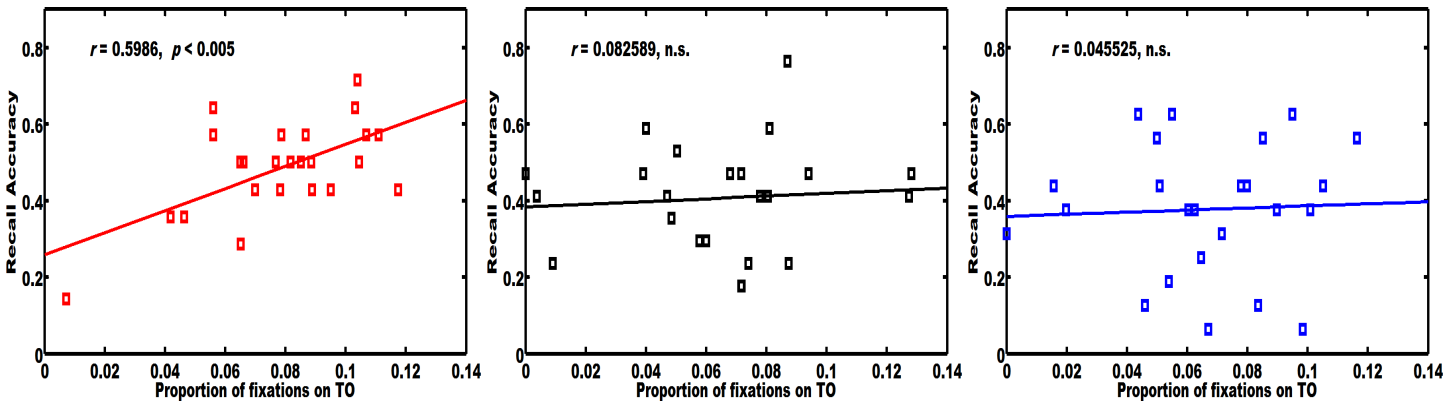
Figure 4: Mean proportion of correct answers corresponding to the different emotion types for IM-related questions. Error bars denote one standard deviation.

Linking fixations on target objects to memory

One of the main findings of studies with static scenes is that people better remember the items that they have fixated (Melcher & Kowler, 2001; Hollingworth, 2006). To further explore the link between where people look and what they remember when watching

movies in terms of specific objects, we examined the proportion of eye-fixations that fell on target objects (TOs) that were tested in the memory questions. We checked, for each participant, if more fixations on the TOs meant a better recall accuracy for each of the emotion categories. Specifically, we computed the mean proportion of fixations fixated on all TOs for each emotion category, and the mean accuracy for the questions for that category of emotion. Finding a strong correlation between proportion of fixations on the target objects, and proportion of correct answers in the memory test would imply a strong one-to-one correspondence between fixating an object and remembering it. Alternatively, this relationship might be weaker in the case of dynamic movie stimuli. Participants might have made use of peripheral vision to encode object properties. Moreover, some of the location-questions involved second order relationships. For example, to know that the flower vase is to the left or right of the bed, ideally, the viewer should have known the locations of both the flower vase and the bed.

To begin with, we computed the absolute time for which the TOs were visible on screen for the different emotional stimuli– the visibility duration of TOs denoted as a fraction of the clip length was found to be $0.24 \pm 0.18, 0.16 \pm 0.22$ and 0.13 ± 0.24 for the neutral, positive and negative stimuli respectively. Right-tailed two-sample t-tests revealed that neutral TOs appeared for marginally longer than negative TOs ($t_{28} = 1.9052, p = 0.0674$), but were not visible for significantly longer than positive TOs ($t_{29} = 1.089$, n.s.). The visibility durations of positive and negative TOs were very comparable ($t_{29} = -0.4518$, n.s.). We then computed Pearson correlations between the proportion of fixations on TOs corresponding to IM questions for a particular emotion, and the proportion of correct answers for those questions (Figure 5). While a significant positive correlation was observed for neutral movie clips, as would be expected from previous studies of static images, this relationship broke down for positive and negative valence clips. On one hand, replication of the link between where people look and what they remember, previously reported with static images, for neutral movie stimuli provides confirmation that results with photographs can also hold with more complex, naturalistic viewing conditions. However, failure to find such a strong correlation for emotional stimuli calls into question the generalization of the idea purporting a close link between fixation and memory. Thus, participants might fixate an object but fail to remember it or, conversely, remember items that were not directly fixated when viewing emotional content.



(a) (b) (c)

Figure 5: Pearson correlation between proportion of fixations on the target objects and proportion of correct answers for questions corresponding to (a) neutral, (b) positive and (c) negative stimuli. Correlation is computed based on the data for 24 subjects.

By also including questions about spoken dialogue, we were able to examine memory for non-visual (and non-fixated) aspects of the movie stimuli. Memory performance was better for the questions regarding the spoken dialogue (up to 78% correct) compared to the visual details (accuracy ranging between 48% - 53% for the three question types). Although it is not possible to directly compare these two types of questions, the results indicate that participants were paying close attention to the spoken dialogue, but any difference between visual and auditory questions was not likely due to a ceiling effect. For dialogue-related questions, a one-way ANOVA test failed to reveal any effect of emotion ($F_{(2,12)} = 0.67$, n.s.), and this was confirmed by post-hoc t-tests ($t_{12} = -0.8322$, n.s.) comparing

memory performance for neutral vs emotional clips. Thus, our findings regarding the links between emotion, eye movement patterns and visual objects were not likely to be the result of some non-specific influence of emotion on all types of memory tests.

Analysing the memory performance for visual detail-related questions produced contrasting results as compared to the case where we considered visual and dialogue-related questions together. A two-way ANOVA showed that in this case, there was no main effect of 'question type' ($F_{(2,46)} = 1.5$, n.s) or interaction effects ($F_{(4,46)} = 1.52$, n.s), while the influence of 'emotion category' on memory performance was marginally significant ($F_{(2,46)} = 3.11, p = 0.0561$). When the comparison was repeated with the data for positive and negative valence clips pooled together, the F-statistic reached significance ($F_{(1,46)} = 5.8, p < 0.05$). This finding was reinforced through post-hoc t-tests- there was a marginally significant difference between the memory performance for neutral and emotional (positive-cum-negative valence) clips ($t_{45} = 1.8, p = 0.0784$). Also, the memory performance for visual details differed significantly between neutral and negative clips ($t_{28} = -2.08, p < 0.05$), while there was no performance difference between neutral and positive ($t_{29} = 1.29$, n.s.) or negative and positive valence clips ($t_{31} = 0.48$, n.s.). These findings are suggestive of a specific effect of emotion on where people look in the scene and how this relates to what they remember, rather than a more broad influence on memory *per se*.

Influence of low-level and high-level factors on memory

We then investigated which aspects of the stimuli contributed to poorer memory for details of emotional stimuli as compared to neutral stimuli. In addition to emotional content, the movie clips varied with respect to a number of attributes. As described previously, we identified seven possible factors which could have influenced memory performance. These included relatively low-level factors, such as clip length and number of shots, as well as the time at which, and duration for which, target objects appeared on screen. In principle, these factors could have had orthogonal effects on performance (if they did not differ systematically between the different emotion categories) or confounding effects if they varied along with emotional category– a correlation analysis between the low-level factors and clip valence revealed only a significant moderate correlation with L_c ($\rho = -0.3195, p < 0.05$). We performed a series of analyses to determine the influence of stimuli-related low and high level factors on memory performance for visual details (see previous section for description). The high-level factors included duration of emotional content as well as the mean emotional valence and arousal ratings.

The partial correlations between memory recall accuracy (RA), and the aforementioned factors is as outlined in Table 2. The correlation coefficient and corresponding p value are respectively listed in parentheses. From the table, it can be inferred that (i) The clip-length and number of shots in the movie clip least influence recall accuracy. (ii) Maximum and significant positive correlation is observed between recall accuracy and the visibility duration of target objects, as expected. (iii) Participants better recalled details of those objects that were mainly visible in the initial part of the stimulus, indicating a primacy effect. (iv) Only a weak correlation was observed between high-level factors-LE_c, V_c, A_c and memory performance. Nevertheless, it needs to be noted here that these correlations only capture linear relationships between the variables, and the recall accuracy-valence relationship is not linear as evident from the above discussion.

Table 2: Partial correlations for the considered factors and IM performance for visual questions. Correlation coefficients (ρ) and p values are respectively listed in parenthesis.

	L_c	VD_t	TD_t	LE_c	NS_c	V_c	A_c
RA	$(-0.1055, 0.5155)$	$(0.3121, 0.047)$	$(0.2969, 0.0594)$	$(-0.1739, 0.277)$	$(-0.031, 0.8474)$	$(-0.233, 0.1426)$	$(0.2479, 0.1181)$

Therefore, we performed a backward linear regression analysis in order to determine which set of predictors best explained the observed recall accuracy. Table 3 presents the results. When the model included only a pair of variables, the clip valence and arousal scores provided the best prediction, accounting for 9.3% of the observed variance in recall accuracy, while clip length and the number of shots were least predictive. When combined with valence and arousal factors, the visibility duration (VD_t) turned out to be the most predictive among the low-level factors. Finally, the best-fit linear model comprised both high and low-level factors ($V_c, A_c, VD_t, TD_t, LE_c$),

Table 3: Backward linear regression analyses with various predictor combinations. R^2,$\overline{R}^2$ denote raw and adjusted coefficient of determination. F denotes the F-statistic, and p is the observed significance level.

Model variables	R^2	$\overline{R}^2$	F	p
L_c, NS_c	0.005	-0.0399	0.1165	0.8903
VD_t, TD_t	0.029	-0.0146	0.6692	0.5173
V_c, A_c	0.0933	0.052	2.265	0.1158
V_c, A_c, LE_c	0.093	0.0302	1.477	0.2342
V_c, A_c, TD_t	0.1014	0.0388	1.619	0.199
V_c, A_c, VD_t	0.144	0.0843	2.411	0.0799
V_c, A_c, VD_t, LE_c	0.1489	0.0678	1.837	0.1397
V_c, A_c, VD_t, TD_t	0.1969	0.1205	2.754	0.051
V_c,A_c,VD_t,TD_t,LE_c	**0.2338**	**0.1404**	**2.502**	**0.0457**
$V_c, A_c, VD_t, TD_t, LE_c, L_c, NS_c$	0.2431	0.1072	1.789	0.1171

and accounted for 23.4% of the observed variance in memory performance, with a significant F-statistic. Here again, it is imperative to note that much of the unaccounted variance in memory performance may have resulted from factors such as (i) varying difficulty of the IM questions, as some questions could have been presumably harder than others and (ii) familiarity of participants with the presented movie clips, which could have influenced the observed results.

Discussion

The first finding of this experiment was that fixations were more focused and constrained for emotional clips compared to neutral ones. As reported for eye movement patterns with movie trailers (Dorr et al., 2010), there was less spread in eye movements within participants, and more agreement between participants when viewing an emotional movie. Moreover, these eye movement patterns were quite different from those typically found in studies of eye movements for static scenes, in which participants often scan a large portion of the image (as also found in our study with a static image control condition described below). When watching movies, participants seemed to focus their fixations on the most important aspect of the movie frame for the ongoing narrative and avoided making many exploratory saccades. In addition, fixation durations were relatively long compared to those found with reading or natural scenes, as has previously been reported (Dorr et al., 2010; Smith & Mital, 2013). Overall, the pattern of eye movements for movies seems to differ in many ways from the classic findings reported with static scenes (Buswell, 1935; Yarbus, 1967).

We also examined the relationship between where participants looked and what they remembered. As described above, there was a tendency to focus gaze on specific objects central to the narrative in emotional scenes, which might have hurt memory performance for questions about details of peripheral objects. Consistent with this hypothesis, memory was particularly poor for questions about object details in emotional movie clips. In the case of negative valence movie clips, for example, participants performed around chance level on questions about the location of specific objects in the scene. In contrast, memory for the details of the auditory conversation, which would have been more central to the events taking place in the clip, remained good for both neutral and emotional movies. While the expected relationship between where participants looked and what they remembered was replicated for neutral movie clips, it was not found for the emotional clips. In principle, this lack of correlation could be caused by failure to remember fixated items or, conversely, an ability to remember details for non-fixated items (such as by using a wider span of attention). In support of the first interpretation, a

recent study has shown that task constraints affect whether or not fixated objects are encoded into memory (Tatler & Tatler, 2013).

One interpretation of these results is that the emotional content tended to draw more attention (Kensinger et al., 2007), at the expense of processing specific object details that were peripheral to the plot of the movie. If so, gist of the scenes, rather than peripheral object details, would have been encoded better in memory. To test this hypothesis, we tested participants 8 weeks later on their ability to recognize whether they had seen a clip previously or not. This recognition task, unlike the object memory test for the first experiment, could benefit from stronger memory for the gist of the movie clip content.

Experiment 2- Long-term memory (LTM) for scene gist

Procedure

Only 13 of the 24 participants returned for the 'old/new' recognition test scheduled eight weeks after the original experiment. They were presented with a total of 49 clips (clip length μ = 6 s, σ = 1.2 s) which included ('old') snippets of the 30 clips shown previously, plus 19 clips from the same movies and with the same actors, but not seen in the original experiment. Of the 19 'new' snippets, 10 were similar in content to the emotional clips seen in the main study (5 similar in content to the negative valence clips and 5 similar to the positive valence clips), while the other 9 clips were extracted from emotionally neutral movie portions.

Results

To investigate if emotional valence again influenced observers' visual behavior while viewing snippets in the LTM test, we computed values of the previously considered eye movement variables from the LTM data. Since no emotional ratings were acquired for the 'new' clips shown in the LTM test (even if they were visually similar to the original neutral/emotional clips as stated earlier), forthcoming comparisons between the individual emotion categories will pertain only to the 30 'old' snippets extracted from movie clips, while comparisons between neutral and emotional conditions will involve all the 49 snippets used.

As for movie clips, highest fixations per second were observed for positive stimuli, while highest fixation durations and largest saccade amplitudes were observed for neutral stimuli. No main effect of emotion was revealed by 1-way ANOVA comparisons of per-second fixations and fixation durations, even though post-hoc t-tests showed significant differences between fixations per second for neutral vs negative ($t_{12} = 2.424, p < 0.05$) and positive vs negative ($t_{12} = 4.4173, p < 0.001$), as well as between fixation durations for positive vs negative ($t_{12} = -3.0578, p < 0.001$) and neutral vs positive ($t_{12} = -2.5859, p < 0.05$) snippets. ANOVA comparison of saccade amplitudes however, revealed a main effect of emotion ($F_{2,38} = 8.48, p < 0.001$) with post-hoc t-tests confirming the differences between neutral and negative ($t_{12} = -5.6292, p < 0.0005$), and neutral and positive ($t_{12} = 4.1786, p < 0.0005$) as significant. In contrast to the IM test, highest entropy was observed for positive stimuli and the main effect of emotion on entropy differences was revealed by an one-way ANOVA test ($F_{2,38} = 4.64, p < 0.05$), with paired t-tests showing entropy differences between neutral vs positive ($t_{12} = -2.3787, p < 0.05$), neutral vs negative ($t_{12} = -2.5132, p < 0.05$) . and positive vs negative ($t_{12} = -7.6435, p < 0.00001$) as significant. Upon extending these analyses to also include the 'new' clips, we observed that only saccade amplitude differences between neutral and emotional stimuli remained significant ($t_{12} = -3.9234, p < 0.005$).

Concerning long-term memory recall, participants were able to correctly recognize the 30 previously seen clips. However, this recognition performance was much better for emotional clips. While 59.2% of the previously viewed neutral clips were classified as 'old' on an average, 80.4% of 'old' emotional clips were recognized as having been seen before (80% for positive valence clips and 81% for negative valence clips). A post-hoc t-test revealed that the effect of (positive or negative) emotion on hit rate was significant ($t_{28} = -3.057, p < 0.005$). For the 19 clips that were not part of the original experiment, participants were able to correctly reject the 'new' clips. Still, fewer emotional clips were rejected (53.1% correct rejection) as compared to neutral clips (69.2% correct rejection) even though the difference in correct rejections was not significant ($t_{17} = 1.1578$, n.s.). Further analyses to investigate if the emotional valence had any influence on the tendency to reject a 'new' clip revealed interesting trends- participants were more adept at rejecting

the 'new' positive valence clips as compared to negative valence clips- post-hoc t-tests showed that the difference in correct rejections was significant between neutral and negative clips ($t_{12} = 2.18, p = 0.0499$), but not between neutral and positive clips ($t_{12} = -0.1109$, n.s.).

Given the hit and rejection rates in the 'old/new' recognition test, we further investigated if emotion influenced the sensitivity of participants using signal detection theory analysis. The mean sensitivity ($\bar{d'}$) for the neutral, positive and negative stimuli were found to be 0.8789,1.4317 and 0.5492 respectively, implying best detection performance for positive valence clips. Also, the criterion bias values indicated a conservative bias for neutral clips ($\bar{C}_{neu} = 0.1317$), in contrast to a liberal bias for positive ($\bar{C}_{pos} = -0.1446$) and negative ($\bar{C}_{neg} = -0.666$) valence clips. A 1-way ANOVA test confirmed the main effect of emotion on both sensitivity ($F_{(2,38)} = 12.18, p < 0.0001$) and criterion bias ($F_{(2,38)} = 8.8, p < 0.001$) of participants.

Finally, we attempted to find if there were any eye movement differences between the snippets that a participant recognized or rejected. To this end, we considered 'emotion type' (neutral/emotional) and 'decision type' (accept/reject) as two factors, and performed two-way repeated measures ANOVA tests for the aforementioned eye movement variables. Interestingly, a few significant differences showed up– comparison of saccade amplitudes revealed the significant main effect of decision type ($F_{(1,51)} = 5.68, p < 0.05$) and a marginal interaction effect ($F_{(1,51)} = 3.73, p = 0.0595$) , with paired t-tests indicating significant differences only between accepted and rejected emotional snippets ($t_{12} = 2.3233, p < 0.05$). On the other hand, entropy comparisons revealed only a significant effect of emotion type ($F_{(1,51)} = 5.08, p < 0.05$) with paired t-tests confirming that entropy was significantly higher for rejected neutral snippets (3.237) than emotional snippets (3.0824) ($t_{12} = -3.1409, p < 0.01$). Given the specificity of these differences, we can only conclude that eye movement patterns *per se* cannot predict whether a stimulus will be recognized or rejected in the scene-gist recognition test.

Discussion

The visual behavior of participants while viewing movie clips and 'old' snippets was found to be similar in a number of ways– however, only a significant difference in saccade amplitude between neutral and emotional clips remained when the eye movement analysis was extended to include the 'new' snippets.

Consistent with our prediction on long-term emotional memory, participants tended to recognize emotional movie clips better than the neutral ones. The pattern of hits and false alarms, however, differed across the three different emotion conditions. Lower sensitivity and the most conservative criterion were observed for neutral valence clips, implying that participants often failed to recognize 'old' neutral clips and consistently rejected 'new' neutral stimuli. Positive clips had the highest $\bar{d'}$, but also a less stringent criterion than neutral clips. Negative clips, however, corresponded to the most liberal criterion and the least sensitivity, suggesting that participants found negative clips familiar even when they had not actually seen them before. This finding is consistent with the suggestion of Rimmele and colleagues (Rimmele et al., 2011), that emotion can enhance the subjective feeling of familiarity. Nevertheless, we could not identify any correlation between participants' visual and recall behavior as all the considered eye movement factors were found to incapable of predicting whether a given stimulus will be recognized/rejected by the observer.

Experiment 3- Eye movements and memory for matched static scenes

Procedure

In order to directly compare eye movements and memory for static scenes and movies, an additional condition was run for the participants immediately after the long-term memory experiment. We chose 23 of the 60 questions for which accuracy in the visual scene detail questions was less than 50% (participants had to typically select one out of three choices for these questions)- 6, 8 and 9 questions respectively corresponded to the neutral, positive and negative valence clips. For these questions, we extracted one key-frame from the clip that contained the target object(s). These key-frames were then shown to participants for the same duration as the target objects had been visible in the original experiment. However, when the target objects were visible for long durations, the key-frames

were shown for a maximum of 10 seconds. To ensure direct comparison to the earlier study with movie clips, identical questions were asked with the static scenes. We then compared the proportion of correct answers provided by each subject for the control and original experiments respectively.

While the key-frames were extracted from both neutral and emotional movie clips, the individual frames themselves were neutral in valence since the emotional content of the movies depended on narrative elements. This was confirmed by ratings obtained from an independent group of eight subjects who were not part of any of the experiments. The participants rated each of the static frames for valence on the same five-point scale (-2 to 2) used previously in the first experiment. Average ratings were near zero for the static frames from each emotional valence category. Paired t-tests failed to reveal any significant difference in valence for the key-frames corresponding to different emotions. Results comparing eye movements and memory performance for static and dynamic scenes are reported based on the data available for 12 subjects who participated in both the original and control experiments.

Results

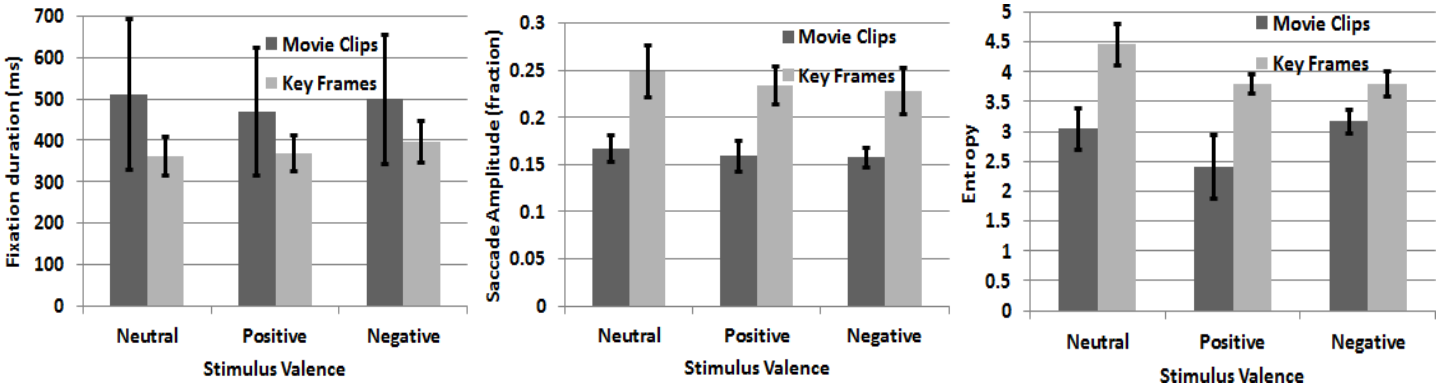


Figure 6: Comparison between (a) eye fixation durations, (b) saccade amplitudes and (c) fixation entropy for movie clips vs static images.

Comparing eye movements for static scenes against those for the corresponding video shot(s) from the original movie clips, we found a significant difference between the eye movement patterns for static and dynamic scenes in line with previous studies. On average, participants made shorter fixations, longer saccades and explored a larger area of the screen (Figure 6). This finding was consistent with other recent studies comparing static and dynamic viewing conditions (Dorr et al., 2010; Smith & Mital, 2013). Two way repeated measures ANOVA tests with 'emotion type' and 'stimulus type' as factors for the fixation duration, saccade amplitude and entropy variables revealed the following– only the main effect of stimulus type ($F_{(1,71)} = 16.98, p < 0.000001$) was identified for differences in fixation duration. Both emotion ($F_{(2,71)} = 3.79, p < 0.05$) and stimulus type ($F_{(1,71)} = 260.28, p < 0.000001$) accounted for saccade amplitude differences. For entropy differences, the effect of emotion ($F_{(2,71)} = 23.43, p < 0.000001$) and stimulus type ($F_{(1,71)} = 221.03, p < 0.000001$) as well as interaction effects ($F_{(2,71)} = 11.23, p < 0.0001$) were revealed to be significant.

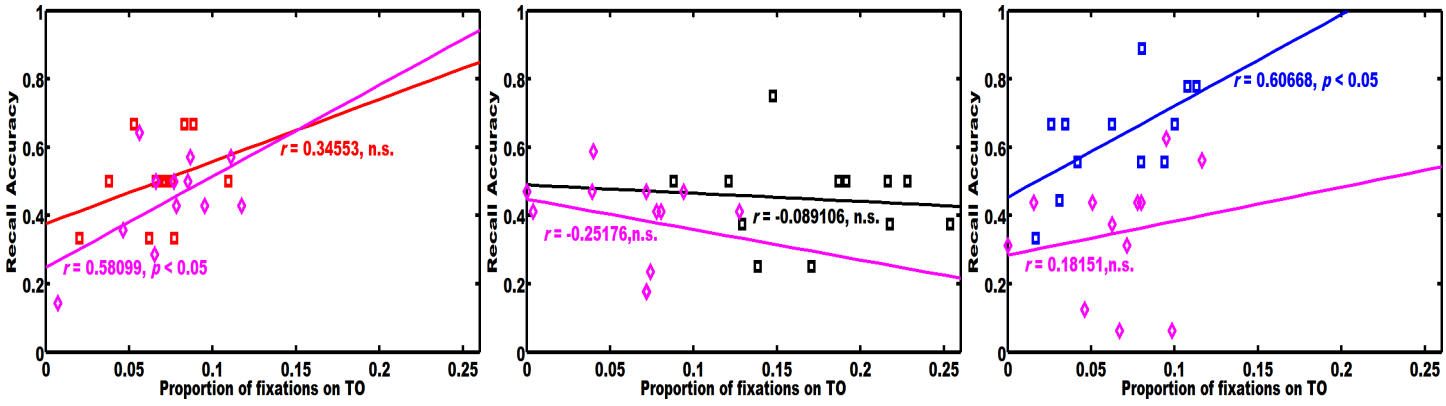
Considering the different emotion categories, post-hoc paired t-tests revealed that eye fixation durations on key-frames were significantly different from those on corresponding video shots for neutral ($t_{11} = 3.0791, p < 0.05$), positive ($t_{11} = 2.421, p < 0.05$) and negative ($t_{11} = 2.4626, p < 0.05$) valence stimuli. Similarly, saccade amplitudes were also significantly different for the neutral ($t_{11} = -7.5276, p < 0.00005$), positive ($t_{11} = -8.6849, p < 0.000005$) and negative ($t_{11} = -10.6987, p < 0.000001$) emotion categories. Comparing entropy, which measures the dispersion of fixations across the static image/video shot, significant differences were again observed for the neutral ($t_{11} = -7.9596, p < 0.00001$), positive ($t_{11} = -8.5608, p < 0.000005$) and negative ($t_{11} = -7.4496, p < 0.00005$) emotions. In addition to the differences between static frames and movies, we also compared per-

formance on static images corresponding to the different emotion categories. Fixation durations were slightly longer for negative key frames, as confirmed by paired t-tests for negative versus neutral ($t_{11} = 3.5185, p < 0.005$) and positive frames ($t_{11} = 2.684, p < 0.05$), while there was no significant difference in saccade amplitudes.

Fixation entropy was higher for neutral key-frames, as confirmed by t-tests between neutral vs positive ($t_{11} = -6.3404, p < 0.0001$) as well as neutral vs negative key-frames ($t_{11} = -8.1131, p < 0.00001$). It is interesting to note, however, that the pattern of results was different for the key frames and the original movie clips. For example, mean fixation duration was maximum for neutral movie clips, but minimum for the corresponding key-frames. Thus, there were important differences in eye movement patterns between movie clips and static key-frames.

Correlating fixations with memory, we found that the close relationship between where people look and their memory for object details was replicated only for neutral clips in the IM experiment. We checked if there was any difference in the proportion of fixations on target objects between the IM and control experiments. Our analysis revealed that the proportion duration for which the TOs were fixated was higher in the static control experiment. Paired t-tests revealed the difference was significant for neutral ($t_{11} = 3.7385, p < 0.005$) and positive stimuli ($t_{11} = 9.7266, p < 0.0000005$), while being marginally significant for negative stimuli ($t_{11} = 1.6005, p = 0.0689$). This confirms that participants were able to look around more broadly at potential target objects (on whom the IM questions were based) in the case of static images compared to dynamic movie scenes.

We then investigated whether the increased proportion of fixations on TOs in the control experiment resulted in better memory for details pertaining to those TOs (Figure 7). There was a positive correlation between the fixation duration and recall accuracy for neutral stimuli as in the original experiment, but this correlation was not significant. The correlation for positive stimuli was near zero, but in contrast to the original experiment, a strong positive correlation was found for negative valence stimuli. Thus, in the case of negative static images, where participants looked and what they remembered was closely linked.



(a) (b) (c)

Figure 7: Pearson correlation between proportion of fixations on the target objects in key-frames, and proportion of correct answers for questions from (a) neutral, (b) positive and (c) negative emotion categories. Correlation is computed based on the data for 12 subjects. Purple data points and lines correspond to performance of the 12 control subjects in the IM test (Figure 5 presents the corresponding plot). Figure is best viewed in color.

Furthermore, participants were significantly better with recalling details of static images (53.6% accuracy) compared to their earlier performance with the movie clips (36.2% accuracy) from which the key frame had been chosen. A two-way repeated measures ANOVA with 'emotion type' and 'stimulus type' as factors to compare the memory performance for movie clips vs static images, showed the main effects of emotion type ($F_{(2,71)} = 5.43, p < 0.01$) and stimulus type ($F_{(1,71)} = 11.31, p < 0.005$), as well as interaction effects ($F_{(2,71)} = 8.91, p < 0.0005$).

Paired t-tests reinforced this finding– the difference in memory performance between movie clips and static images was significant ($t_{11} = -3.5669, p < 0.005$). This difference, however, was mediated by the emotional content of the scene. Specifically, there

was no difference for neutral movies vs neutral static images ($t_{11} = 0.7609$, n.s.). In contrast, for negative stimuli, the difference in memory performance between clips and images was highly significant ($t_{11} = -5.4905, p < 0.001$), while the difference was marginally significant for positive valence stimuli ($t_{11} = -2.1117, p = 0.058$).

As the IM and control experiments involved different number of subjects, we also analyzed the performance of the 12 control subjects in the IM experiment to verify if aforementioned differences were indeed due to stimulus/valence-related factors instead of statistical power. Purple data points in Figure 7 denote performance of the control subjects considering all visual questions in the IM test (subset of the data points plotted in Figure 5), and the estimated linear correlations are also shown. Similar to Figure 5, a significant linear correlation is still observed between the proportion of TO fixations and recall accuracy for the control subjects considering neutral stimuli, even though the significance level is higher in this case. Also, as previously, the corresponding correlations for positive and negative valence stimuli is not significant.

Since the same amount of visual information available for inferring scene details in the IM and control experiments as movie clips and key-frames were matched with respect to TO visibility durations, the observed results provide further evidence that emotional content reduced memory (and also visual attention) for object details in movies. The specificity of this effect, limited to emotional clips, also argues against the potential confound of an order effect (the static image test was run eight weeks after the original movies had been shown), which might have been expected if participants were just better in general at the task eight weeks later due to repetition of the stimuli, repetition of the questions or an overall improvement in performance in the task with practice. Instead, these results suggest that negative emotion, in particular, reduces visual memory for scene details in movies.

Discussion

Here, we directly compared where people look and what they remember for static and dynamic images from Hollywood movies, as a function of the emotional content of the movie clips. As expected, participants looked around more at the static image compared to a dynamic movie, covering more of the potentially interesting (in terms of the memory questions) areas of the scene. Comparing eye movement patterns for key-frames vs corresponding video shots from the movie clips, we can confirm the influence of emotion on memory, consistent with the idea that emotion, in particular negative emotion, focuses the viewer on the most emotionally-charged aspect of the scene with little or no resources left for encoding more peripheral details (Kensinger, 2009). While this effect was most pronounced for negative valence movies, positive movies also showed a similar reduction in performance compared to corresponding static key-frames. These results provide further evidence that emotion has specific effects on memory for peripheral object details.

It is interesting to note that memory performance and some eye movement parameters differed across emotion categories even for the static scenes. Given that the key-frames extracted from the original movie clips were relatively neutral in emotion, as described above, the difference in fixation durations and entropy between emotion categories is somewhat surprising. One possible explanation is that there was some influence of previous exposure to the original movie clip. If so, then it is likely that such influences were mainly implicit, since $\bar{d}'$ for negative clips was not better than for neutral ones. Even for the positive movie clips, participants were not particularly good ($\bar{d}' = 1.43$) at discriminating them from new clips. The key-frames was much shorter than the clips used in the recognition test, so performance for recognizing whether static key frames were seen previously would be expected to be even worse than for the short clips. Moreover, the static frames were typically not extracted from the most emotional part of the movie clip. It is important to note that eye movement patterns were different for the key frames and the original movie clips. For example, mean fixation duration was maximum for neutral movie clips, but minimum for the corresponding key-frames. Thus, the main result of this experiment was that there were large differences in eye movement patterns between movie clips and static key-frames.

Finally, memory for object details in neutral movies and in static images was quite similar. We had expected that the simple fact that participants were watching movies would reduce encoding of peripheral details. One possibility is that the fact the participants anticipated an upcoming test about the details of the scene counteracted any tendency, which might otherwise have been present, to focus only on the narrative aspects of the visual scene. Likewise, the fact that movie clips often involved camera motion could have reduced

memory for location and for objects that disappeared and reappeared over time. Other factors, however, might have been expected to result in better scene memory for movies. For example, the increased ecological validity of the movie stimulus could have made the dynamic scene more easy to remember (Tatler & Melcher, 2007). Also, the fact that static images were from previously viewed scenes (from the IM Experiment) could have led to some overall benefit in this experiment. Thus, our current results do not allow us to make a definitive statement about differences in detailed memory for static versus dynamic scenes. Instead, our results can only provide further evidence for a specific effect of emotion on memory for scene details compared to scene gist.

General Discussion

Overall, our results are consistent with the hypothesis that emotional content in a movie clip increases the tendency of participants to focus their gaze on the central theme of the scene rather than more peripheral details. There are four converging lines of evidence for this idea. First, the pattern of gaze was more focused for emotional movie clips compared to neutral ones. Second, immediate memory for the details of movie clips was worse for emotional movies. Third, scenes with emotional content were recognized more often during a long-term memory test, consistent with the idea that the gist of the scene was encoded into long-term memory. Finally, there was a specific difference in memory performance for movies versus static scenes found only for emotional movies. Together, these findings all argue for a role of emotion in focusing attention and gaze to only the most emotionally salient aspects of the movie and improving gist memory at the cost of memory for more peripheral visual details.

In general, eye movements while watching movies differed from what would have been expected from classic studies using static scenes (Buswell, 1935; Yarbus, 1967). As noted previously (Tosi et al., 1997; Goldstein et al., 2007; Dorr et al., 2012; Wang et al., 2012; Smith & Mital, 2013), there was less spread in eye movements within participants and more agreement between participants when viewing movies, in particular when viewing emotional movies. Fixation durations were also longer than those typically found during reading or viewing of static scenes (Dorr et al., 2010; Smith & Mital, 2013). The current findings suggest that some of these differences are mediated by the emotional content of movies. When interpreting the finding that participants keep their eyes near the center of a movie trailer for example (Dorr et al., 2010), this may reflect relatively low-level processes like a high rate of motion and placing of narrative elements at the center of the image (for review, see Smith, 2013), but it may also be due in part to the highly emotional nature of movie trailers.

Another interesting aspect of the present results is the specific pattern of sensitivity and bias found for neutral, positive and negative movie clips in the long-term recognition test. Judgments of both positive and negative emotion clips showed a more liberal criterion compared to neutral clips. This finding is consistent with previous suggestions that emotional valence may enhance the subjective impression of familiarity without necessarily improving sensitivity (Rimmele et al., 2011). Our results suggest that emotion might promote false memories, making it more likely for someone to think that they remember something based on familiarity alone. There was, however, a difference between positive and negative movie clips in terms of sensitivity. Compared to neutral clips, participants were more sensitive to positive clips (suggesting enhanced long-term memory) and less sensitive to negative clips. It has long been noted that there is a bias to remember positive, rather than negative, events over time (Kensinger, 2009). Here, we found this effect after only 8 weeks with memory for movies, rather than autobiographical memories of personal events. Also, our finding that this influence on long-term recognition was valence-specific argues against a purely arousal-based explanation of the role of emotion on memory(as in Sharot & Phelps, 2004) for movies. Instead, emotional scenes in movies seem to be remembered differently than neutral parts of the movie, and this difference is specific to whether the scene is positive or negative in emotion.

More generally, the current results show the value in using movie stimuli to bridge the gap between photographs and real life visual experience. While movies differ from our day-to-day experiences in important ways, they still capture some important aspects of real-world vision which takes place in a dynamic context and often involves narrative elements and emotional content. A good movie guides the viewer's attention and gaze to the most salient aspects of the audio-visual stimulus in order to follow the story. Instead, most studies of picture viewing have allowed participants to look around the scene for memory or search tasks. While we do often search for

objects in real-life, most of our eye fixations are tied to particular tasks such as guiding actions, conversing with others or working (for review, see Tatler et al., 2011). The dynamic nature of movies makes gaze control even more important than for static scenes, since there is no extra time to look around the scene since the eye must follow the most important aspects of the scene in order to track unfolding events. Recent studies (Vig, Dorr, Martinetz, & Barth, 2011) have demonstrated the anticipatory nature of eye movements while viewing realistic dynamic scenes due to our inherent knowledge of the surrounding world, and emotional objects are likely to play an active role in gaze guidance. Conversely, the current results suggest that the relationship between where people look and what they remember may vary depending on the stimulus, *e.g.*, a naturalistic movie with no edits conveying a neutral emotion (used in some previous studies) vis-à-vis an emotional Hollywood movie.

Moreover, movies involve scenes and events that unfold over time and allow participants to create predictions and more lasting event representations. One of the important challenges for vision science is to study active perception and not just reactive responses to unpredictable, briefly presented stimuli. In the case of movies, the visual system must take advantage of the predictability of real-world scenes to match objects and people across saccadic eye movements and frequent cuts in order to understand the meaning of events. Future studies could take advantage of these aspects to understand the way in which visual processing is influenced by contextual factors, such as a stable environment and the ability to create predictions, during event perception.

Conclusion

In summary, the pattern of results found here are consistent with the idea that movies use narrative and clever editing techniques to direct the gaze of the viewer to the most relevant aspects of the scene, at the expense of processing of and memory for more peripheral details. Similarly, real-world tasks can also focus visual processing on the most task-relevant aspects of the scene (for review, see Tatler et al., 2011). The addition of emotion seems to further enhance this focus, leading to strong memories for the gist of the event but at the cost of poor memory for peripheral details. Thus, the gaze patterns and impressive memory capacity for natural scenes shown in previous studies, while valuable in demonstrating the abilities of human visual memory, may not always be used in dynamic, task-defined situations in which visual processing and memory are focused only one or a few items at each point in time.

Acknowledgements

Ramanathan Subramanian was supported by A*STAR Singapore under the Human Sixth Sense Program (HSSP) grant. Nicu Sebe was supported by the FP7 European Union project xLiMe, and David Melcher was supported by a European Research Council (ERC) grant (grant agreement no. 313658).

References

Attar, C. H., Andersen, S. K., & Müller, M. M. (2010). Time course of affective bias in visual attention: Convergent evidence from steady-state visual evoked potentials and behavioral data. *NeuroImage*, *53*(4), 1326–1333.

Bartolini, E. E. (2011). *Eliciting emotion with film: Development of a stimulus set*. Honors theses - all, Wesleyan University.

Buchanan, T. W., & Adolphs, R. (2002). The role of the human amygdala in emotional modulation of long-term declarative memory. In S. Moore & M. Oaksford (Eds.), *Emotional cognition: From brain to behavior.* London, UK: John Benjamins.

Buswell, G. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.

Calvo, M. G., & Lang, P. J. (2004). Gaze Patterns When Looking at Emotional Pictures: Motivationally Biased Attention. *Motivation and Emotion*, *28*(3), 221–243.

Dorr, M., Martinetz, T., Gegenfurtner, K., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*, 1–17.

Dorr, M., Vig, E., & Barth, E. (2012). Eye movement prediction and variability on natural video data sets. *Visual Cognition*, *20*, 495–514.

Furman, O., Dorfman, N., Hasson, U., Davachi, L., & Dudai, Y. (2007). They saw a movie: long-term memory for an extended audiovisual narrative. *Learning & memory*, *14*(6), 457–467.

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine*, *37*(7), 957–964.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of films. *Projections: The Journal for Movies and Mind*, *2*(1), 1–26.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, *14*(1), 40–48.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.

Hirose, Y., Tatler, B., & Regan, O. (2010). Perception and memory across viewpoint changes in moving images. *Journal of Vision*, *10*, 1–20.

Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual research. *Visual Cognition*, *14*, 781–807.

Hollingworth, A., Williams, C., & Henderson, J. (2001). To see and remember: visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, *8*(4), 761–768.

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. In *Iccv.*

Kaspar, K., Hloucal, T.-M. M., Kriz, J., Canzler, S., Gameiro, R. R. R., Krapp, V., et al. (2013). Emotions' impact on viewing behavior under natural conditions. *PloS one*, *8*(1).

Kensinger, E. A. (2009). Remembering the Details : Effects of Emotion. *Emotion review*, *1*(2), 99–113.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*(4), 575–591.

Melcher, D. (2010). Accumulating and remembering the details of neutral and emotional natural scenes. *Perception*, *39*.

Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, *41*(25–26), 3597–3611.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8).

Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision*, *9*(10), 2.1–2.12.

Rimmele, U., Davachi, L., Petrov, R., Dougal, S., & Phelps, E. a. (2011). Emotion enhances the subjective feeling of remembering, despite lower accuracy for contextual details. *Emotion*, *11*(3), 553–562.

Sharot, T., & Phelps, E. A. (2004). How arousal modulates memory: disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(3).

Smith, T. J. (2013). Watching you watch movies: Using eye tracking to inform cognitive film theory. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring cognition at the movies.* New York: Oxford University Press.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*(8), 1–24.

Soleymani, M., Pantic, M., & Pun, T. (2012). Multimodal emotion recognition in response to videos. *IEEE Transactions on Affective Computing*, *3*(2), 211–223.

Tatler, B., Gilchrist, I., & Land, M. (2005). Visual memory for objects in natural scenes: from fixations to object files. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *58A*(5), 931–960.

Tatler, B., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of vision*, *11*(5), 1–23.

Tatler, B., & Melcher, D. (2007). Pictures in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception*, *36*, 1715–1729.

Tatler, B., & Tatler, S. (2013). The influence of instructions on object memory in a real world setting. *Journal of Vision*, *13*(2), 5.

Tosi, V., Mecacci, L., & Pasquali, E. (1997). Scanning eye movements made when viewing film: preliminary observations. *The International journal of neuroscience*, *92*(1-2), 47–52.

Vig, E., Dorr, M., Martinetz, T., & Barth, E. (2011). Eye movements show optimal average anticipation with natural dynamic scenes. *Cognitive Computation*, *3*(1), 79-88.

Viola, P., & Jones, M. J. (2004, may). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154.

Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, *12*(1), 16.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.