Salience of the lambs: A test of the saliency map hypothesis with pictures of emotive objects

Katherine Humphrey
Division of Psychology, Nottingham Trent University, Nottingham, UK

Geoffrey Underwood
School of Psychology, University of Nottingham, Nottingham, UK

Tony Lambert
Department of Psychology, Faculty of Science, The University of Auckland, Auckland, New Zealand

Humans have an ability to rapidly detect emotive stimuli. However, many emotional objects in a scene are also highly visually salient, which raises the question of how dependent the effects of emotionality are on visual saliency and whether the presence of an emotional object changes the power of a more visually salient object in attracting attention. Participants were shown a set of positive, negative, and neutral pictures and completed recall and recognition memory tests. Eye movement data revealed that visual saliency does influence eye movements, but the effect is reliably reduced when an emotional object is present. Pictures containing negative objects were recognized more accurately and recalled in greater detail, and participants fixated more on negative objects than positive or neutral ones. Initial fixations were more likely to be on emotional objects than more visually salient neutral ones, suggesting that the processing of emotional features occurs at a very early stage of perception.

Keywords: emotion, scanpaths, visual saliency


Introduction

Our ability to rapidly detect emotive stimuli has been studied extensively. It is suggested that the automaticity of processing has evolutionary links, in that the fast and reliable detection of positive and negative reinforcers facilitates adaptive behavior, finally promoting survival and reproductive success (Cacioppo, Gardner, & Berntson, 1999; Lang, Bradley, & Cuthbert, 1997; Öhman, Flykt, & Lundqvist, 2000). Evidence for such rapid detection comes from autonomic measures (electrodermal responses and heart rate), facial EMG responses (corrugator and zygomaticus), startle blink reflex (cf. Bradley, 2000; Hamm, Schupp, & Weike, 2003), and event-related potential studies (Schupp, Junghofer, Weike, & Hamm, 2004). Mogg and Bradley (1999) found that in a dot-probe task, participants were faster to respond to emotional stimuli (positive or negative) than neutral ones, suggesting that they were already attending to the emotional picture before the target dot appeared. In search tasks, advantages have been found for phobic animals among non-phobic distracters (Öhman, Flykt, & Esteves, 2001), but equally search detection advantages have been found for pictures of threatening and pleasant animals (Tipples, Young, Quinlan, Brooks, & Ellis, 2002).

Negativity and positivity effects

Much of the previous literature has found that our attention is automatically drawn to negative information more strongly than it is automatically drawn to positive information (Hansen & Hansen, 1988; Pratt & John, 1991), although emotion-induced enhancements for positive stimuli have also been found (e.g., Schupp, Junghofer, Weike, & Hamm, 2003). The “automatic” attention to negative information has been termed a “negativity effect” (or “negativity hypothesis”), and attention to positive information has been termed a “positivity effect.” For example, Kissler and Keil (2008) found that when instructed to make a saccade toward a picture in the right peripheral visual field, facilitation occurred only for pleasant pictures and saccadic reaction times toward unpleasant pictures were slowed. Similarly, the negativity effect can be seen in event-related potentials (Ito, Larsen, Smith, & Cacioppo, 1998; Smith, Cacioppo, Larsen, & Changrand, 2003), faster reaction to subliminally presented negative stimuli (Dijksterhuis & Aarts, 2003), and increased skin conductance responses even when presented subliminally (30 ms of exposure) and backwardly masked (see Öhman & Mineka, 2001). According to the aforementioned evolutionary theory, the faster detection of negative stimuli may have developed as a survival mechanism.
frequently found negativity effect may occur because negative stimuli such as a deadly poisonous spider require more immediate action (and thus faster processing) than, for example, a cute cuddly kitten, which poses little threat. In this sense, evaluation of threat may be a key underlying component of this evolutionary system (Öhman, 1993; Öhman et al., 2001). Öhman (1993) suggested that threat information is first processed by a feature detection system, which “tags” the stimuli as ancestrally or behaviorally relevant and passes the information to the organism’s arousal system, which optimizes selective attention and orienting. Individuals then compare the stimuli with earlier memories and respond appropriately. However, many fears are irrational, and although they may have evolutionary relevance, the individual may never have had an actual bad experience with the feared stimulus and thus have no associated negative memories to compare it to. Despite this criticism, evidence for rapid processing of fear stimuli has been found. Fear-relevant targets among neutral distracters are detected faster than neutral targets among fear-relevant distracters (Öhman et al., 2001). However, this was only true if the target was feared by that individual—no effect was present for generic fear-relevant stimuli if they were not feared by the viewer. This raises the question of why there are so many individual differences in feared stimuli—if this system is based on evolutionary dangers, everyone should hold the same fears. Furthermore, these individual differences cannot be completely explained by fears developed through bad experience because many phobias are irrational.

The emotionality effect

Eye movement analysis has also provided evidence for an advantage of emotional stimuli irrespective of whether it is negative or positive. For example, Alpers (2008) found a left visual field bias regardless of valence and Lang, Greenwald, Bradley, and Hamm (1993) found that increased viewing duration for both pleasant and unpleasant scenes relative to neutral scenes. This equal advantage for positive and negative stimuli over neutral stimuli is referred to as the “emotionality effect” or “emotionality hypothesis” and is an alternative to the “negativity hypothesis.” Calvo and Lang (2004) found that when emotional and neutral pictures were presented simultaneously, the probability of placement of the first fixation and the proportion of viewing time during the first 500 ms were higher for both pleasant and unpleasant pictures than for neutral pictures, suggesting emotional meaning captures initial overt orienting and engages attention early.

The emotionality effect has also been found when pictures are presented in the periphery (Nummenmaa, Hyona, & Calvo, 2006), suggesting that the processing of emotional stimuli occurs at a very early stage, even before the eyes move to focus on the target. Nummenmaa et al. found that the probability of first fixating an emotional picture and the frequency of subsequent fixations were greater than those for neutral pictures. Even when participants were instructed to avoid looking at the emotional pictures, first fixations were still more likely to fall on emotional stimuli than neutral stimuli.

Emotional gist

It could be argued that in studies such as the aforementioned Calvo and Lang (2004) experiment, participants may have used emotional gist to process the stimuli top-down, as each emotional picture was an entirely positive or negative scene rather than a scene with a positive or negative region of interest. Gist is an interesting concept, as previous studies have implied an emotional gist just by changing an object in the scene. The current paper is different in that a “negative” stimulus will contain a target with a negative valence, but the presence of this object will not change the overall gist of the scene (i.e., the scene remains neutral). There are many examples of different studies that have implied/defined “gist” in different ways. For the simplicity of this paper, “gist” is referred to as the global (overall) emotionality, whereby the scene can be rapidly identified as positive, negative, or neutral without having to explore the individual (local) features of the scene. For example, a picture of the aftermath of a bomb explosion would hold a negative gist, whereby every scene feature would be negative and related to the overall “story,” e.g., dead bodies, blood, broken windows, rubble, building remains, fire—there is not just one negative feature but an overall negative emotional valence. The current research uses complex neutral background scenes, e.g., suburban streets, and a single emotional feature/region, e.g., a dead animal. The importance of the viewer exploring the scene is twofold. First, it allows us to test whether a single emotional feature does, in fact, attract eye fixations over other areas in the picture, and second, it allows us to test whether emotional saliency has an overriding effect of more visually salient areas (which have previously been shown to attract fixations in free-viewing tasks, e.g., Foulsham & Underwood, 2008).

The current paper aims to test this rapid processing theory suggested by Nummenmaa et al. (2006), through analysis of the first fixation location, to see whether it still holds true when emotional gist is not obvious but instead requires deeper processing. In the real world, it is more common for a neutral scene to contain an emotive object (e.g., a spider in the bedroom) than a scene to hold an overall positive or negative valence, e.g., a bomb explosion or car crash scene.

Visual saliency

It is often the case that emotional objects in a scene are also visually salient, take for example Figure 1.
saliency is how much something stands out from its neighboring items, based on color, intensity, and orientation change. The lamb in Figure 1 holds positive valence, but when analyzed using a saliency algorithm (Itti & Koch, 2000), the lamb is also the most visually salient part of the picture. If one thinks about many fear-inducing stimuli (negative valence), they also hold high visual saliency, for example, a black spider on a white wall. Is it therefore the emotional valence that draws our attention toward these objects, or is it actually due to their low-level visual characteristics (i.e., saliency)? In the current research, we try to separate these factors by exploring whether emotional items are still fixated when in competition with other more visually salient items in the picture.

It has previously been shown that saliency guides our eye movements when inspecting a scene (Humphrey & Underwood, 2009; Underwood & Foulsham, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006), and saliency maps have even been used to predict which image observers were viewing, given only scanpath data (Cerf, Harel, Huth, Einhauser, & Koch, 2008). However, the influence of saliency is moderated depending on the task (e.g., Einhäuser, Rutishauser, & Koch, 2008; Underwood, Templeman, Lamming, & Foulsham, 2008) and the top-down knowledge of the viewer (Humphrey & Underwood, 2009). If low-level visual saliency is an important factor in attracting attention, then this should still be true when the most visually salient object is not the most emotionally salient one. However, if valence plays a special role in this attraction, then it could result in a sort of cognitive override similar to the effect of domain-specific knowledge/expertise. In a way, emotional valence (e.g., the recognition or experience of emotion) is something that most cognitively unimpaired human beings are experts in. Evidence for the influence of emotional valence includes increased fixation on unpleasant than on neutral stimuli (Christianson, Loftus, Hoffman, & Loftus, 1991) and when presented in the left visual field negative stimuli, an increased “leftward bias” (a tendency for people to first look into the left visual field), whereas if presented in the right visual field, leftward bias is decreased (LaBar, Mesulam, Gitelman, & Weintraub, 2000).

Emotion, visual saliency, and memory

There is also evidence that both emotional valence and visual saliency affect memory for a target or scene. For example, Sheth and Shimojo (2001) briefly displayed a target and then asked participants to point to its previous location. Participants estimated targets to be closer to the center of gaze and closer to visually salient markers in the visual display than they actually were. The locations of objects presented earlier were remembered falsely as being closer to visually salient reference frames than they really were. On the other hand, increased amygdala activity at encoding for negative words is significantly correlated with enhanced recognition memory for negative words tested after a short delay (Hamann & Mao, 2002). Furthermore, patients with amygdala lesions fail to show normal emotional enhancement of memory for positive and negative emotional pictures (Cahill, Babinsky, Markowitsch, & McGaugh, 1995; Hamann, Lee, & Adolphs, 1999).

The memory advantage for emotional regions of interest in a picture (especially negative ones) may also be accompanied by diminished memory for other details in the picture, a phenomena known as attentional narrowing. An example of this is the weapon focus effect, whereby the presence of a weapon captures attention at the expense of memory for other details such as the perpetrator’s facial characteristics and clothing. When participants were asked to identify a target from a line up, they were less accurate if the target had previously held a gun compared to if they had previously held a checkbook (Loftus, Loftus, & Messo, 1987). Eye tracking analysis showed that while watching the original slideshow containing the target person, participants fixated longer and more frequently on the gun than the checkbook. Similarly, Hope and Wright (2007) found that if the critical object was a gun, participants were more accurate and detailed in their descriptions of the object, but recognition accuracy of the person was impaired. Participants were also less confident when questioned about the target person. The memory advantage for the critical object cannot be simply explained by increased fixations upon it, as the effect was still found when participants were only allowed to make one fixation (Christianson et al., 1991). Another aim of the current
experiment is to see if there is any evidence of attentional narrowing. It is predicted that participants will accurately recall fewer neutral peripheral details of the scene when an emotionally negative region of interest exists.

Previous research has shown that emotion can act as a top-down influence on eye movements, attention, and memory. However, colorful or bright objects also have this power, and many previous studies have not completely controlled for visual saliency, although some low-level factors directly related to perceptual saliency have been previously regulated. Calvo and Lang (2004) controlled for luminance alone; Nummenmaa et al. (2006) controlled for RMS contrast and color saturation; and Calvo, Nummenmaa, and Hyönä (2008) and Nummenmaa, Hyönä, and Calvo (2009) also controlled for energy, kurtosis, and skewness. However, the current study measures saliency using Itti and Koch’s (2000) saliency map algorithm, which is more comprehensive, taking into account orientation, intensity, and color. This paper also controls for both visual and emotional saliency to establish which has a greater influence on overt attention (eye movements) and on memory.

**Methods**

**Participants**

Twenty-four participants with normal or corrected-to-normal vision were recruited through an opportunity sample (participants available at the time) of the University population.

**Materials and apparatus**

Eye movements were recorded using a remote Eyegaze eye tracker, which uses corneal reflection to determine gaze direction, sampling at 60 Hz. No attachments to the head were required. Advanced image processing algorithms in the Eyegaze System explicitly accommodate several common sources of gaze point tracking error, including head range variation and pupil diameter variation (typical average bias error of 0.45 degree).

A set of high-resolution digital photographs were created by “creating” real-life scenes of neutral settings (e.g., a bedroom) containing target objects that held a positive, negative, or neutral valence. For as much as possible, the targets were really in the scenes when the pictures were taken; however, in order to get enough variation and range of valences, some digital editing was required to add target objects to the scenes. These images were run through Itti and Koch’s (2000) saliency program, to identify the first five most visually salient regions for each picture. Measures of saliency were based on changes in color, intensity, and orientation. Standard parameters were used, as far as possible, for example, the default setting for the FOA is a size equivalent to 1/16th of the image, which it is argued is a realistic estimate of the resolution of human visual attention. The pictures were edited so that the most visually salient feature was not the most emotionally salient one. The emotionally salient object and the most visually salient object were always equidistance from the central fixation cross. The distance between the most visually salient object and each of the three emotional objects (positive, neutral, and negative) was also balanced.

A pilot study whereby participants were shown 120 stimuli and asked to judge whether any of the pictures looked unrealistic or “strange” was carried out. It was emphasized that some pictures may have objects in strange places, but it was the physical characteristics rather than the location of these objects they were to judge. This was to make sure that any photo editing had not compromised the realism of the scenes. Second, the participants were asked to rate how positively, negatively, or neutrally emotive the pictures were, using the same rating scales as those used by Lang et al. (1997). As the stimuli were all, in effect, neutral scenes containing either a positive, negative, or neutral target object, but without an overall emotional gist, participants were asked to inspect the scenes carefully and note down any items that stood out and whether those items made the participants feel happy or unhappy and excited/aroused or calm. Participants were given two 9-point scales, one measuring valence of the pictures, ranging from very happy (1) to very unhappy (9), and the other measuring arousal of the pictures, ranging from very excited (1) to very calm (9). From this pilot with 15 participants, 90 pictures that were unambiguously rated as positive (average valence 2.1; average arousal 3.1), negative (average valence 7.9; average arousal 2.7), or neutral (average valence 4.9; average arousal 5.1) were chosen, with 30 pictures in each category. Thirty of these pictures (10 positive, 10 negative, and 10 neutral) were randomly selected to appear at the encoding stage of the experiment, and the remaining 60 were fillers at a later recognition test. For all the stimuli (at encoding and recognition), the ratings for emotional saliency were balanced between positive pictures and negative pictures. The neutral pictures contained a neutral object the same size and distance from the center as the emotional object in the positive and negative pictures (see Figures 2a–2c). Positively emotive Regions of Interest (RoIs) included domesticated animals (cats, kittens, puppies, etc.), positively rated food (cakes, chocolate, sweets, cookies), and infant wildlife (e.g., fawn). Negative RoIs included gory or unpleasant objects (e.g., roadkill), animals shown to provoke high fear ratings (e.g., snakes and spiders), and weapons (e.g., knives, guns etc.). Neutral control objects included everyday items but in unusual places, e.g., a handbag in the road, a glove on a park railing, a kettle in the living room, etc. The neutral control target was not the
The most visually salient region of the picture. Each emotive (or control) RoI appeared on an otherwise neutral background, including suburban streets, interiors (living room, kitchen, bedroom, etc.), educational settings (e.g., classroom), or parks. As previous studies have found that the presence of a person in a picture attracts eye fixations (Humphrey & Underwood, 2010), it was decided that people should not be included in the stimuli.

The emotionally neutral (and non-visually salient) object was included in neutral pictures to try to match them as closely as possible with the emotional pictures. If the neutral picture was simply a “still” of, for example, a countryside view, nothing is happening or has happened—there is no “story.” The emotional picture, however, might be taken as a cue to a story about, for example, an animal that was killed in an accident. A difference in eye movements may then be due to the participant trying to understand what happened in the emotive scene (e.g., fixation of the cars to look for evidence that one of the cars hit the animal). Previous research (Christianson & Loftus, 1991) has, in addition to the neutral (a woman riding on a bicycle) and emotional (a woman lying beside her bicycle wounded) scenes, included an “unusual condition” (a woman carrying a bicycle on her shoulder). Participants in the “unusual” condition performed differently from those in the emotional condition, suggesting that emotional valence does affect performance beyond the emotional feature being “unusual.”

The average distance of the positive, negative, or control (neutral) RoI from the center was also balanced over the three categories of stimuli, as was its size. Pictures were displayed at a resolution of 1600 by 1200 pixels. The authors intended to measure scanpath similarity using a string editing technique (e.g., Foulsham & Underwood, 2008), which involves dividing the picture up into a 5 × 5 grid. As such, the stimuli were designed so that the RoIs (emotionally salient and visually salient regions) fell within one of the equally sized 25 regions of the picture. The RoIs were on average 220 pixels, with a range of 200 to 240 pixels, and visually salient, and emotive RoIs were matched for size to the nearest 10 pixels. Visually salient RoIs were also matched for intensity of salience across the stimulus categories. Both visually salient and emotive RoIs were objects, or parts thereof, rather than empty spaces (e.g., the sky). Appendix A outlines the emotive and salient objects for each of the 30 pictures shown at encoding.

At the recognition test, the original 30 pictures were designated as “old” (shown at both encoding and recognition) and the remaining 60 were labeled “new” (shown

Figure 2. (a) An example of an emotionally positive stimulus. The emotional area of interest here is the cat on the bed. The most visually salient object being the bottom left picture in the picture frame above the cat. (b) An example of an emotionally negative stimulus. The emotional area of interest here is the dead animal on the road. The most visually salient region being the top of the tree in the right-hand corner. (c) An example of an emotionally neutral stimulus. The picture contains an emotionally neutral yet unexpected object (kettle on the rug). The most visually salient region is the picture frame on the windowsill (left).
only as fillers at test). New pictures were created using a similar neutral background setting as the old pictures, e.g., from a different angle or a different room in the house but were from the same categories (suburban streets, interiors, etc.). The similarities/differences between the old and the new items were equivalent for the negative, neutral, and positive scenes, so that there were no noticeable differences in complexity either between old and new scenes or between stimulus types. Therefore, any recognition differences are not due to the negative old/new scenes being more different than the neutral old/new scenes, for example. The old and new stimuli contained similar emotive target objects (i.e., domesticated animals, positively rated food, infant wildlife, gory or unpleasant objects, fear-provoking animals, weapons, and everyday items in unusual places) and were also matched for saliency. Old and new stimuli were also matched for RoI distance from the central fixation point.

**Procedure**

Participants were informed that they would see a series of pictures and they were to remember as much detail as possible in preparation for a memory test. They were not told anything about the emotional nature of the pictures. Participants were not told that their eye movements were being recorded but instead lead to believe that their pupil size was being monitored. Following a 9-point calibration procedure, and a practice using unrelated pictures, participants were shown thirty stimuli (10 positive, 10 negative, and 10 neutral) in a randomized order. Each picture was preceded by a fixation cross and was presented for 3 s.

After all 30 pictures had been presented, participants performed a distracter task (computer game) for 10 min and were then given 10 min to complete a recall memory test. They were instructed to write down five things they remembered about each picture, in order of importance. Participants were given an “answer sheet” that simply numbers 1 to 30 and had a five-line space for them to write any details they remembered. It was emphasized that they did not have to recall the pictures in the order that they were presented.

Immediately after the recall test, the distracter task was performed for a further 10 min. Participants were then shown a second set of pictures and had to decide whether each picture was old (from the first set of pictures) or new (never seen before). Ninety pictures were presented in a random order; 30 of these were old and 60 were new. Each picture was again shown for 3 s and participants could only make a response after this time.

**Results**

Two main types of data were collected, memory data (recognition and recall) and eye tracking measures—the average number of fixations in visually salient and emotive Regions of Interest (RoIs), the average fixation duration in RoIs, and the position of the first fixation. Trials were excluded where the fixation at picture onset was not central or when calibration failed. Data attrition was less than 2%.

**Recognition accuracy (old/new judgments)**

The accuracy of correctly recognizing old pictures as old (“hits”) was calculated. Negative pictures were correctly identified as old 92.5% of the time (hit rate 0.925); neutral pictures were correctly identified as old 87.91% of the time (hit rate 0.879); and positive pictures were correctly identified as old 89.58% of the time (hit rate 0.896). A repeated measures ANOVA found a reliable difference in accuracy between the picture types, $F(2,46) = 6.790, MSE = 19.022, p < 0.01, \eta^2 = 0.043$. Post-hoc tests revealed more accurate recognition for negative over neutral stimuli, $t(23) = 3.817, SEM = 1.20072, p < 0.01$, and no difference between positive and neutral stimuli or between negative and positive stimuli (see Figure 3).

**Recall memory**

The average number of references to visually salient RoIs in the written responses was calculated for negative, neutral, and positive pictures. On average (across persons), participants recalled at least one detail about each of the thirty pictures, although this was not necessarily a detail about either visually salient or emotional RoIs (e.g., vague detail about the background scene, “kitchen”). The number of points recalled ranged from zero to five per picture. The ANOVA found a main effect of picture type, $F(2,46) = 742.543, MSE = 0.010, p < 0.001, \eta^2 = 0.944$, with post-hoc tests finding more references to visually salient areas in neutral rather than negative emotional pictures, $t(23) = 27.024, SEM = 0.0369, p < 0.001$, and positive emotional pictures, $t(23) = 31.099, SEM = 0.0311,$
t-tests found reliable differences between visually salient and emotive RoIs for positive stimuli: \( t(23) = 30.753, SEM = 0.3184, p < 0.001 \); negative stimuli: \( t(23) = 31.978, SEM = 0.03544; p < 0.001 \); and neutral stimuli: \( t(23) = 12.541, SEM = 0.6412, p < 0.001 \). For positive and negative stimuli, more references were made to emotive RoIs than visually salient RoIs, but for neutral pictures, more references were made to visually salient RoIs than emotive RoIs.

### Average number of Fixations in RoIs (per picture)

Eye movement data were analyzed. An eye movement was classified as a saccade when its velocity reached 30 deg/s or when its acceleration reached 8000 deg/s². A fixation was defined as anything above 70 ms—microfixations below 70 ms were discarded. On average, participants made 10 fixations per stimulus (an average of 100 fixations over the 10 stimuli in each emotive category, e.g., 100 fixations over all 10 negative stimuli, etc.) An ANOVA compared the average number of fixations per picture that fell inside the visually salient RoIs, finding a difference between the picture types, \( F(2,46) = 43.063, p < 0.001, \eta^2 = 0.456 \). Comparisons found that there were on average more fixations on visually salient RoIs in neutral rather than emotionally negative pictures, \( t(23) = 8.071, SEM = 0.4652, p < 0.001 \), and also more fixations in these areas of neutral rather than positive pictures, \( t(23) = 7.999, SEM = 0.5105, p < 0.001 \). There was no reliable difference between negative and positive pictures, \( t < 1 \) (see Figure 5a).

A t-test was carried out to compare the average number of fixations per picture that fell in emotive RoIs on positive, negative, and neutral pictures (Figure 5b) and found that more fixations fell in negative RoIs than on positive RoIs, \( t(23) = 2.107, SEM = 0.17798, p < 0.05 \); more fixations fell in negative RoIs than neutral RoIs (neutral objects in unusual places), \( t(23) = 6.06, SEM = 0.14509, p < 0.05 \); and more fixations fell in positive RoIs than neutral RoIs, \( t(23) = 2.622, SEM = 0.19229, p < 0.05 \).

The average number of fixations that fell in visually salient RoIs and emotive RoIs was compared. For both negative and positive pictures, reliably more fixations were made in emotive RoIs than in visually salient RoIs: \( t(23) = 8.181, SEM = 1.504, p < 0.05 \) and \( t(23) = 6.174, SEM = 1.369, p < 0.05 \), respectively. For neutral pictures, more fixations fell in visually salient RoIs than emotive RoIs (neutral objects in unusual locations): \( t(23) = 7.603, SEM = 0.057, p < 0.05 \).

### First fixation

The location of the first fixation on each stimulus was identified and the percentage of these that fell inside visually salient and emotive RoIs was calculated for each...
picture type. There was a main effect of picture type, $F(2,46) = 4.965$, $p < 0.05$, $\eta^2 = 0.101$, with post-hoc indicating that a greater percentage of first fixations landed in visually salient RoIs on neutral pictures than on both positive pictures, $t(23) = 2.429$, $SEM = 2.401$, $p < 0.05$, and negative pictures, $t(23) = 3.444$, $SEM = 1.694$, $p < 0.01$. There was no reliable difference in the number of first fixations to positive and negative stimuli, $t < 1$ (see Figure 6a).

A $t$-test comparing the percentage of first fixations that fell in emotive RoIs on positive and negative pictures found no reliable difference, $t(23) = 1.00$. However, there were differences between negative and neutral pictures, $t(23) = 2.379$, $SEM = 2.80183$, $p < 0.05$, and between positive and neutral pictures: $t(23) = 3.181$, $SEM = 2.88152$, $p < 0.05$ (see Figure 6b).

The average percentage of first fixations to visually salient RoIs was compared to the average percentage of first fixations to emotive RoIs. Paired-samples $t$-tests revealed reliable differences for negative stimuli, $t(23) = 5.042$, $SEM = 3.471$, $p < 0.001$; for positive stimuli, $t(23) = 3.091$, $SEM = 4.853$, $p < 0.01$; and for neutral stimuli, $t(23) = 2.106$, $SEM = 3.16609$, $p < 0.05$.

**Scanpath similarities**

Scanpaths are sequences of fixations and indicate the transitions between fixation locations. This string editing technique (see Underwood, Foulsham, & Humphrey, 2009) involves turning a sequence of fixations into a string of characters by segregating the stimulus into labeled regions. The similarity between two strings is then computed by calculating the minimum number of editing steps required to turn one into the other. Three types of operations are permitted: insertions, deletions, and substitutions. Similarity is given by one minus the number of edits required,
standardized over the length of the string. An algorithm for calculating the minimum editing cost is given in Brandt and Stark (1997; Appendix B) and was adapted to analyze string similarity using a Java program (an example of the Java code used can be found in Appendix C).

Comparisons using t-tests found that scanpaths when viewing both negative and positive pictures were less similar to saliency scanpaths than when viewing neutral pictures: $t(23) = 2.106, SEM = 0.01085, p < 0.05$, and $t(23) = 3.216, SEM = 0.00750, p < 0.01$, respectively. The saliency model made better predictions about scanpaths with neutral pictures than with emotional pictures of either type (see Figures 7a and 7b).

![Figure 7](image)

Figure 7. (a) Bar chart showing the similarity of scanpaths at encoding compared to model-predicted saliency scanpaths, for each stimulus type. Error bars represent standard error of the mean ($SEM$). (b) (Left) An emotionally neutral picture with the saliency-predicted scanpath (red and yellow) and an actual participant scanpath (blue). (Right) An emotionally negative picture with the saliency-predicted scanpath (red and yellow) and an actual participant scanpath (blue). Note the decreased number of fixations in visually salient areas and the multiple fixations on the negative RoI.

Discussion

This experiment aimed to determine how eye movements and memory are affected when emotional saliency and low-level visual saliency are in direct competition. Furthermore, it aimed to find out whether evidence for the emotionality or negativity hypothesis exists. The results suggest that low-level visual saliency does have an effect on eye movements when inspecting a visual scene, but this can be overridden by “emotional saliency” (especially negative emotion). A negativity effect was found for recognition and recall memory and in the percentage of fixations in Rols. In contrast, a general emotionality effect was found for the location of the first fixation, implying that the processing of emotional features occurs at a very early stage of perception.

Eye data analysis revealed a greater number of fixations to negative Rols than to positive Rols and reflects the negativity effect shown in the recall memory data (better recall for negatively emotive details). This pattern of eye movements suggests a difficulty in disengaging attention from negative stimuli (e.g., Gerdes, Alpers, & Pauli, 2008), in that the participant refixates the negative object rather than fixating once and then
exploring the rest of the scene. It could be that looking at threat is a beneficial strategy aimed at (a) monitoring the fear-relevant stimulus and (b) increasing the probability of appropriate responses in case the threat situation changes (Clark, 1999). Participants also made a greater number of fixations to emotive ROIs (positive or negative) than to control ROIs in neutral pictures (neutral objects in unusual locations). This suggests that eye movements were drawn toward negative/positive ROIs due to their emotive nature rather than because they were simply unexpected, for example.

Previous studies have shown that low-level visual saliency helps guide our eye movements in free-viewing tasks, with highest saliency regions being fixated first. If one thinks about commonly feared stimuli, such as spiders, it becomes apparent that these stimuli are also generally highly visually salient (e.g., black spider on a white wall). So, could this be that an attributing factor to feared or negative stimuli attracting our attention? How much of this “automatic processing” is due to its emotive nature and how much to its low-level visual characteristics? The current experiment explored this question by presenting scenes with both visually salient regions and separate emotionally salient regions. An increased number of written references (during recall) to visually salient ROIs in neutral pictures was accompanied by a reliably higher percentage of fixations in visually salient ROIs and a decreased number of fixations in “control” ROIs (neutral objects in unusual locations), demonstrating that visual saliency does guide our attention. However, when emotive ROIs were also present, the increased number of written references to these areas was accompanied by a reliable increase in the percentage of fixations to emotive ROIs at the expense of fixations to visually salient ROIs. The attentional capture by these emotive areas could be interpreted as a cognitive override of saliency, similar to that seen with domain expertise (Humphrey & Underwood, 2009). It seems as though our attention to emotive objects in a scene (e.g., a spider) cannot be purely due to their low-level visual characteristics but is also influenced by the valence of such stimuli. This is supported by the differences in scanpath similarities between picture types. Scanpaths produced when viewing neutral pictures were reliably more similar to model-predicted saliency scanpaths than when viewing negative or positive pictures, suggesting that the presence of emotional ROIs attracts attention more than the visually salient regions.

Additionally, analysis of the position of first fixation revealed that the presence of the emotive ROI (positive or negative) decreased the likelihood of first fixating a visually salient ROI. This “emotionality” effect suggests an automatic processing of emotional stimuli, irrespective of whether they hold positive or negative valence, and supports previous findings that difficulty in disengaging [an emotive stimulus] is not necessarily accompanied by an initial orienting bias (Calvo & Lang, 2004; Nummenmaa et al., 2006). Furthermore, this emotionality effect (e.g., Alpers, 2008; Calvo & Lang, 2004; Lang et al., 1993) of positive and negative ROIs on eye movements is still present when the overall gist of the scene is neutral and deeper processing is required.

The findings of this study also indicate that memory can also be affected by emotional content of a scene. At recognition, participants had to decide whether they had seen a picture before or not. The results show a reliable memory advantage for negative pictures, which were remembered more accurately than neutral pictures. Although there was no reliable difference in accuracy between negative and positive pictures, indicating that recognition memory was also high for positive pictures (thus suggesting an emotionality effect), these positive stimuli did not have the advantage over neutral pictures that negative stimuli did. Overall, the results suggest a negativity effect on recognition memory.

A recall memory test, where participants had to write down five things they could remember about each picture, in order of importance, was also carried out. The results show that reliably more references were made to visually salient Regions of Interest on neutral stimuli than on positive or negative stimuli. When an emotive ROI existed, the number of references to visually salient regions significantly decreased. The increase in recall for emotional details at the expense of recall for other details could be interpreted as evidence of attentional narrowing (Easterbrook, 1959). However, this did not affect recognition memory for emotional pictures, as reported in some previous experiments (Hope & Wright, 2007; Loftus et al., 1987). The memory advantage for emotional ROIs was reliably greater for negative stimuli than positive stimuli, suggesting a negativity effect.

The advantage for negative pictures at recognition and recall tests could be due to a tendency for the viewer to “personalize” the emotive stimuli at encoding, relating them to memories of similar emotional events that individuals may have experienced in the past (Heuer & Reisberg, 1990). The emotional memory may then serve as a cue at retrieval (Kensinger & Corkin, 2003). The advantage of negative over positive stimuli could be due to the infrequency of negative events relative to positive ones, and thus, negative memories may be more specific and distinct. For example, one may see cute kittens every other day but have only ever been bitten by a spider once, thus making the negative memory more distinct. The negatively emotive stimulus that the participant relates to that memory may, therefore, be more accurately recognized or recalled at a later time. It should also be acknowledged that as the recognition test was preceded by a recall test, there is the possibility that performance in the recognition test could have been affected by the prior recall test. Future research could rerun the experiment using either the recall or recognition test in isolation or counterbalance the order of these tasks over participants.
It has previously been suggested that informing participants (for ethical reasons) that they will be presented with emotional pictures may bias them to deliberately search for emotive ROIs (Calvo & Lang, 2005). However, emotive stimuli in the current study were not predictable. First, positive, negative, and neutral stimuli were presented in a random order, and second, the ROIs were in unique locations in every picture. Emotive ROIs were also (on average) equally distanced from the central fixation point.

In summary, low-level visual saliency does have an effect on eye movements when inspecting a visual scene, but this can be overridden by “emotional saliency” (especially negative emotion). Attentional narrowing for negative regions of interest was demonstrated in the recall memory test. A negativity effect was found for recognition and recall memory and in the percentage of fixations in ROIs. In contrast, a general emotionality effect was found for the location of the first fixation, implying that the processing of emotional features occurs at a very early stage of perception. It is suggested that participants have difficulty in disengaging their attention from negative stimuli, which might be a defense mechanism to fear- or threat-related material.

Appendix A

List of objects in areas of interest of the three types of stimuli

The types of scenes were balanced over the stimulus types, by using different variations of the same category, e.g., negative, positive, and neutral scenes would all include a “kitchen” background, but each kitchen would be different (Table A1).

<table>
<thead>
<tr>
<th>Stimulus type</th>
<th>Scene</th>
<th>Emotive ROI</th>
<th>Salient ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Lounge/living room</td>
<td>Decapitated mouse</td>
<td>Window</td>
</tr>
<tr>
<td>Negative</td>
<td>Kitchen</td>
<td>Vomit</td>
<td>Kettle</td>
</tr>
<tr>
<td>Negative</td>
<td>Bedroom</td>
<td>Bloody knife</td>
<td>Bag</td>
</tr>
<tr>
<td>Negative</td>
<td>Park</td>
<td>Gun</td>
<td>Swing</td>
</tr>
<tr>
<td>Negative</td>
<td>Bathroom</td>
<td>Feces (diarrhea)</td>
<td>Taps</td>
</tr>
<tr>
<td>Negative</td>
<td>Garage</td>
<td>Spider in web</td>
<td>Lawn mower</td>
</tr>
<tr>
<td>Negative</td>
<td>Garden</td>
<td>Snake</td>
<td>Flowers</td>
</tr>
<tr>
<td>Negative</td>
<td>Forest</td>
<td>Dead and rotting rabbit</td>
<td>Tree</td>
</tr>
<tr>
<td>Negative</td>
<td>Street</td>
<td>Roadkill</td>
<td>Tree</td>
</tr>
<tr>
<td>Negative</td>
<td>Office</td>
<td>Rotting food</td>
<td>Keyboard</td>
</tr>
<tr>
<td>Positive</td>
<td>Lounge/living room</td>
<td>Chocolates</td>
<td>Rug edge</td>
</tr>
<tr>
<td>Positive</td>
<td>Kitchen</td>
<td>Cupcakes</td>
<td>Sink</td>
</tr>
<tr>
<td>Positive</td>
<td>Bedroom</td>
<td>Cat</td>
<td>T-shirt</td>
</tr>
<tr>
<td>Positive</td>
<td>Park</td>
<td>Husky (dog)</td>
<td>Slide</td>
</tr>
<tr>
<td>Positive</td>
<td>Bathroom</td>
<td>Kitten</td>
<td>Shampoo bottle</td>
</tr>
<tr>
<td>Positive</td>
<td>Garage</td>
<td>Mother and baby rabbits</td>
<td>Hi-vis jacket</td>
</tr>
<tr>
<td>Positive</td>
<td>Garden</td>
<td>Puppy</td>
<td>Flowers</td>
</tr>
<tr>
<td>Positive</td>
<td>Forest</td>
<td>Fawn (infant deer)</td>
<td>Signpost</td>
</tr>
<tr>
<td>Positive</td>
<td>Street</td>
<td>Infant fox</td>
<td>Car</td>
</tr>
<tr>
<td>Positive</td>
<td>Office</td>
<td>Jar of sweets</td>
<td>Highlighter pen</td>
</tr>
<tr>
<td>Neutral</td>
<td>Lounge/living room</td>
<td>Kettle</td>
<td>Picture frame</td>
</tr>
<tr>
<td>Neutral</td>
<td>Kitchen</td>
<td>Hair dryer</td>
<td>Notice board</td>
</tr>
<tr>
<td>Neutral</td>
<td>Bedroom</td>
<td>Bike wheel/tire</td>
<td>Cushion</td>
</tr>
<tr>
<td>Neutral</td>
<td>Park</td>
<td>TV</td>
<td>Slide</td>
</tr>
<tr>
<td>Neutral</td>
<td>Bathroom</td>
<td>Laptop</td>
<td>Toilet cleaner</td>
</tr>
<tr>
<td>Neutral</td>
<td>Garage</td>
<td>Sink</td>
<td>Motorbike</td>
</tr>
<tr>
<td>Neutral</td>
<td>Garden</td>
<td>Vacuum cleaner</td>
<td>Flowers</td>
</tr>
<tr>
<td>Neutral</td>
<td>Forest</td>
<td>Alarm clock</td>
<td>Bluebell patch</td>
</tr>
<tr>
<td>Neutral</td>
<td>Street</td>
<td>Arm chair</td>
<td>Car</td>
</tr>
<tr>
<td>Neutral</td>
<td>Office</td>
<td>Bucket and spade</td>
<td>Plant</td>
</tr>
</tbody>
</table>
Appendix B

The algorithm used by Brandt and Stark (1997) to calculate string similarity

Algorithm for String Editing

The two descriptions below complement each other. The first is the actual computer statements in bold; these statements are in the C programming language. The second are comments in English.

compute_distance ()

Compute distance from string 2 to string 1 using D-matrix.

D [first_to_second] [0] [0] = 0;
for (i=1; i<= comp_trun_size [first_file]; i++)

1. Initialize D-matrix to zero; calculate truncation size.

D[first_to_second] [i] [0] = D[first_to_second] [i-1] [0]+cost_of_first_to_null [first_file] [i];

2. Cost of first to null—do this by deleting each character in string 1 one by one; in most this will equal string length.

for (j=1; j<=comp_trun_size [second_file]; j++)
D[first_to_second] [0] [j] = D[first_to_second] [0] [j-1] + cost_of_null_to_second [second_file] [j];

3. Using the D-matrix proceed to calculate distance.

for (i=1; i <= comp_trun_size [first_file]; i++)
for (j=1; j <= comp_trun_size [second_file]; j++)

4. Control scanning from row to row and from top to bottom.

m1 = D[first_to_second] [i-1] [j-1] + cost_of_first_to_null [first_file] [i] [j];
m2 = D[first_to_second] [i-1] [j] + cost_of_first_to_null [first_file] [i];
m3 = D[first_to_second] [i] [j-1] + cost_of_null_to_second [second_file] [j];
D[first_to_second] [i] [j] = m1;
if (m2<m[first_to_second] [i] [j])
D[first_to_second] [i] [j] = m2;
if (m3<m[first_to_second] [i] [j])
D[first_to_second] [i] [j] = m3;

5. This triple computation takes into account the effect of deletions and insertions on shifting the matching of the D-matrix elements and providing the possibility of sidewise matches. Wagner and associates proved this operation to be valid; this extends the discreet dynamic programming algorithm. The algorithm will select the minimum cost taking advantage of possible shifting.

distance = D[first_to_second] [comp_trun_size [first_file]] [comp_trun_size [second_file]];

6. Lowermost right corner of D-matrix will be the minimum total cost of making string 2 identical to string 1.

Appendix C

The Java code used to calculate string editing distance to compare string similarity

Java code for string edit distance

```java
public double stringEditDistance()
{
    // Compute Levenshtein distance
    //*******************************************
    // based on code at http://www.merriampark.com/ld.htm
    int d[] [] = new int[n+1][m+1]; // matrix
    String a; // string from scanpath a
    String b; // string from scanpath b
    int n; // length of scanpath a
    int m; // length of scanpath b
    int i; // iterates through a
    int j; // iterates through b
    char a_i; // ith character of a
    char b_j; // jth character of b
    int cost; // cost
double sd; // for the edit distance
    double sdNorm; // the normalized edit distance
    double sdSim; // the normalized similarity
    // if either string is empty, distance is length of the other
    n = a.length ();
m = b.length ();
    if (n == 0)
    {
        sd = m;
        return sd;
    }
    if (m == 0)
    {
        sd = n;
        return sd;
    }
    d = new int[n+1][m+1];
    // set the first row/column to integers ascending from 1
    for (i = 0; i <= n; i++)
    {
        d[i][0] = i;
    }
    for (j = 0; j <= m; j++)
    {
        d[0][j] = j;
    }
    // loop through the first string
    for (i = 1; i <= n; i++)
    {
        s_i = a.charAt (i - 1);
        // loop through the second string
        for (j = 1; j <= m; j++)
        {
            t_j = b.charAt (j - 1);
            // compare the two characters
            if (s_i == t_j)
            {
                cost = 0;
            }
            else
```
Acknowledgments

Commercial relationships: none.
Corresponding author: Katherine Humphrey.
Email: Katherine.humphrey@ntu.ac.uk.
Address: Division of Psychology, Nottingham Trent University, Chaucer Building, NG1 4BU, UK.

References


