An Eye-Tracking Study of Tourism Photo Stimuli: Image Characteristics and Ethnicity

Abstract

As tourism’s intangibility leads tourism marketers to rely heavily on visuals such as photographic images, the selection of visual stimuli that attract the target audience’s attention is critical. This study used a triangulated approach that included both self-reports and observational eye-tracking data. Australian and Chinese participants were recruited to view a series of photographic tourism images that either depicted high / low arousal activities and natural / built environments. Australian participants fixated more frequently, and for longer durations, than Chinese. Fixation also varied with the image conditions, with the Chinese group having particularly low fixation durations and counts for the low arousal / natural condition. This study fills the void in visual attention research in tourism and presents a novel approach to understanding the appeal of tourism images to potential tourists.

Key words: eye-tracking; tourism activities; arousal; environment
Introduction

Understanding and managing attention is a key determinant of business success (Carrasco 2011; Davenport and Beck 2013), especially as consumers face ever-increasing amounts of information that they cannot possibly process in detail. Both the capacity and information-processing models of attention highlight the central role of physical ad properties in attracting consumers’ attention (Rosbergen, Pieters, and Wedel 1997), and researchers have noted the superiority of photographic images in attention capture (Wedel and Pieters 2012). Tourism offers an excellent platform to understand visual attention, as the intangibility of its products requires abundant use of images (Rakić and Chambers 2010) and the emphasis on visual components such as pictorials shapes markets’ image of the destination (Feighey 2003). This image affects travel choice, satisfaction and behavioral intention, thus attracting academic interest. Research on destination image largely takes a consumer perspective (Stylidis, Belhassen, and Shani 2014), with a focus on what images travelers hold of a destination and how this image is formed. However, little research is conducted from promotion and advertising point of view, especially regarding the relation between promotional strategies and image formation (Michaelidou et al. 2013; Pan 2011) as well as tourism advertising/visuals’ effectiveness in enticing market attention and shaping destination image (Pan 2011).

To understand potential travelers’ attention to tourism visuals, a research design triangulating eye-tracking measures with survey data was deployed to investigate the influence of low- versus high-arousal images and natural versus built images on attention paid to tourism images and to explore whether this influence differs across traveler groups. Two participant groups (Australian and Chinese) viewed pre-tested photographic stimuli as their eye movements
were recorded. The two groups represent two culturally distinct ethnic groups that differ in terms of visual processing behavior (Chua, Boland, and Nisbett 2005) as well as aesthetic and tourism preferences (Dewar, Li, and Davis 2007; Han 2006). Our study is among the first to recognize these differences, and our results enhance understanding of how people perceive tourism images and how Chinese and Australian participants’ perceptive and attention processes differ regarding these images. This study addresses key literature gaps in relation to:

- **Visual attention**: Insights into the impact of pictorials on attention are largely based on evidence outside tourism or using memory measures such as recall. This investigation relies on tourism photographic images since these comprise a large proportion of tourism materials, including blog posts, review web sites, or formal advertising. People generally derive pleasure, excitement, and interest from photographs as well as symbolic meaning (Lang et al. 1993; Lin, Morgan, and Coble 2013), and tourism images are intended to be highly meaningful and rich in context to achieve this purpose. The study’s tourism focus allows us to extend research on cultural differences in how people gather visual information from images with strictly controlled conditions, such as images of faces (e.g., Miellet et al. 2010) and scenes with a clear focal object (e.g., Chua et al. 2005), to the viewing of highly naturalistic and complex images used in tourism promotions.

- **Tourism visuals**: research on tourism visuals has focused largely on topics such as congruency in image representations of a destination, destination image formation, memory, and typologies of tourist- and marketer-generated photographs (e.g., Michaelidou et al. 2013; Pan 2011), with the selection of photographic stimuli for
promotional purposes based primarily on gut feeling (Dewar et al. 2007). Often the same images are displayed to varying markets irrespective of the viewers’ background, preference, and prior knowledge or experience. Given tourism promotion’s increasing reliance on visuals, to optimize visual persuasion marketing practice must be informed by research.

- Methodological innovation: In the tourism field, consumer evaluations are almost universally drawn from self-reports such as questionnaires or interviews. While these approaches are useful, they are limited and also susceptible to potential biases, including recall difficulties. Despite the ocular nature of tourism, few studies have explored visual and other innovative approaches to research (Michaelidou et al. 2013; Rakić and Chambers 2010). This study investigates the role of eye movements in examination of tourism-related photographic images, and the effect of arousal, environment, and ethnicity on any patterns associated with such movements. The study contributes to tourism methodological literature by using eye-tracking as a complement to the more traditional self-report data.

In sum, this study is both theoretically and practically important. Theoretically, it presents inquiries into visual attention to tourism images and extends tourism research by using observational eye-movement data and taking the perspective of tourism advertising. Practically, it offers insights for image selection in tourism promotion and highlights the importance of considering image characteristics and tailoring images to specific markets.
A Quadrant Model of Tourism Images

This section presents literature support for a quadrant model of tourism images to understand visual attention. The model is based on the environment in which a tourism activity takes place and the level of arousal the activity presents—dimensions that are justified by theories of human–environment interactions and environmental aesthetics of landscape preference. Additionally, tourism advertising reflects product differentiation on the basis of these dimensions, and both dimensions are theoretically linked to attention.

Destination Tourism Environment – Natural versus Built

A theoretical foundation of contemporary anthropology is the nature–culture divide, in which culture is defined as a social entity and nature as a bio-physical entity. In Western society, nature and culture are conceptualized as separate and distinct entities. This divide is manifest in the environment (i.e., natural vs. built) and offers a basis for classifying tourism images/activities. At one extreme are tourist attractions that are predominantly nature-based, including mountains, rainforests, and oceans. At the other extreme are constructed or built tourism attractions, such as theme parks, casinos, and museums. Tourist promotions reflect this differentiation between nature-based tourism and urban commercially oriented attractions of a purpose-built kind. This dichotomy is also present in theories of human–environment interaction (Ulrich et al. 1991), as well as studies of environmental aesthetics (e.g., Kaplan, Kaplan, and Wendt 1972; van den Berg, Koole, and van der Wulp 2003) and destination management/image (e.g., Mihalic 2013; Wang and Davidson, 2010). While environmental aesthetics literature suggests a higher likability of natural over built scenes (Kaplan 1995; Lidwell, Holden, and Butler 2010; Ulrich et al. 1991; Sparks and Wang 2014), whether this holds for tourism images is unclear.
Tourism Activities – High vs. Low Arousal

Extensive psychological research demonstrates arousal to be a central construct in consumer behavior (Massara, Liu, and Melara 2010; Mehrabian and Russell 1974; Wakefield and Blodgett 1996). PAD – a widely cited three-dimensional model of emotion reflecting pleasantness, arousal, and dominance (Mehrabian and Russell 1974) – argues that people will respond to physical and social cues in the environment (e.g., Yani-de-Soriano and Foxall 2006). Drawing on the PAD conceptualization, we propose that arousal serves as a measure of the stimulation tourism activities induce. Emotional arousal is highly relevant to the consumption of tourism activities because emotional benefits motivate engagement with tourism. Emotion is recognized as the psychological factor connecting the push and pull factors in decision making. Tourists are pushed by their emotional needs and pulled by the emotional benefits they derive from consuming tourism products (Goossens 2000), and tourists pursue optimal arousal in travel (Iso-Ahola 1982).

Arousal refers to physical and mental alertness varying from high to low and excited to calm (Lane, Chua, and Dolan 1999; Bigné, Andreu, and Gnoth 2005), with the types of tourist activities ranging from relaxing to stimulating. For example, visiting parks and gardens involves less arousal than pursuits such as gaming, nightlife, and adventure activities (e.g., Gyimóthy and Mykletun 2004; Taks et al. 2009). Because arousal results in pleasant experience that is strongly linked to tourist satisfaction and loyalty, understanding tourists’ response to different types of activities is critical (Bigné et al. 2005).
We propose that one way to examine tourism images is through a two-dimensional model of environment type and arousal level as shown in Figure 1. Types of tourism activities are plotted on the model for illustrative purposes.

*Figure 1*

**The Influence of Ethnicity**

A reasonable assumption is that scenes and activities will vary in their attractiveness to tourists from different groups. For instance, the Chinese view nature as a blend of nature and culture, whereas in the West nature is considered to be distinctly separate from civilization (Wen and Ximing 2008). Therefore, the Chinese do not consider isolated wilderness and wildlife attractive but find the integration of landscape and human culture to be a more ideal representation of “nature” (Wen and Ximing 2008). One study found that Chinese are more interested in tourist images representing natural solitude or containing water features whereas Canadian prefer images associated with exotic adventures (Dewar et al. 2007). The cultural root of this response is the traditional theme of *Shanshui* (water and mountains): “the most impressive aesthetic image received from the Chinese ancestors is that nature should have extraordinarily beautiful waters and mountains” (Han 2006, p. 141).

Additionally, activities such as visiting casinos, shopping, and experiencing local nightlife are among the least important to Chinese outbound travelers (e.g., Li, Xu, and Weaver 2009; Mohsin 2008). These activities are expressions of material culture and occur in a manufactured commercial structure, suggesting that Chinese travelers might show low interest in tourism images of these pastimes. Further, Chinese often prefer passive activities such as going to the beach, lazily walking and sightseeing, and boating (Han 2006; Mohsin 2008), while
Westerners appear to be more interested in active pursuits that entail elements of risk, such as hunting and riding (Han 2006; Xu, Morgan, and Song 2009).

**Attention as a Selective Process**

Visual attention is a selective process of allocating our limited mental capacity to some aspects of the visual environment while ignoring other aspects (Carrasco 2011). Greater attention is associated with better cognitive skills such as reading and learning outcomes (Wedel and Pieters 2012). More relevant to business, attention connects awareness and action, and is linked to the decision making chain prior to the decision to act (Wedel and Pieters 2012). Increased visual attention is involved in brand choice (Atalay, Bodur, and Rasolofoarison 2012), can lead to higher choice likelihood (Krajbich, Armel, and Rangel 2010), and is the key to travelers’ processing of messages in marketing materials to select the stimuli of interest.

Overt visual attention is particularly relevant to viewing a complicated scene (Wedel and Pieters 2012) and can be measured by tracking one’s eye movements (Berto, Massaccesi, and Pasini 2008). An eye-tracking study of tourism images is warranted owing to tourism’s uniqueness as being experience-based and intangible. The purchase of a tourism product is risky, because tourists have to make decisions based on abstract images of the promoted product (Djafarova and Andersen 2010; Kotler, Bowen, and Makens 2010), and to address this risk and communicate the intangible, emotional, and non-utilitarian values of the destination experience, marketers rely heavily on imagery representations. Researchers have called for greater sophistication and creativity in advertisers’ choice of techniques and tools to influence tourist opinions (Djafarova and Andersen 2010), highlighting the importance of understanding potential
tourists’ attention to and engagement with tourism images – an awareness that is enabled by eye-tracking methodology.

**Eye Movement Measures of Visual Attention**

Eye tracking captures objective and real-time data about which elements of a specified stimulus people are attending to (Duchowski 2002; Lorigo et al. 2008). A scene within the stimulus that the participant fixates upon is referred to as a fixation that reflects the participant’s specific areas of interest within the stimulus. Commonly used eye-tracking measures are fixation duration, fixation count, and patterns of saccades (the rapid eye movements between fixations, measured as gaze paths). These measures reveal different facets of visual attention, namely how many elements are attended to and for what period of time or the amount of space covered. Since visual information is not attended to during the saccadic eye movements between fixations, people only “see” when they fixate on an object or element within a scene (Hutton and Nolte 2011). Thus, the targets of fixation and the number and duration of fixations can indicate attention focus within a scene and provide information regarding what might be most dominant in the scene. In contrast, saccadic eye movement and gaze path data reflect how the spatial attention system works by revealing the flow of eye gaze from feature to feature, and assist in tracking what is perceived first versus later in the object of interest, reflecting internal shifts of attention within the space of a visual stimulus (Wedel and Pieters 2008, 2012). Thus, images of tourist activities in advertising or destination promotion materials could be decoded as to the extent of their interest to targeted populations.
Research Hypotheses

Corresponding to the proposed quadrant model using environment and arousal as the defining variables, the attention literature shows that environment and arousal influence attention. Attention restoration theory suggests that natural scenes are restorative and can more readily engage attention than scenes of a built environment. Providing opportunity to reflect, viewing natural scenes allows the attention system to rest and recover more than exposure to built environments (Berto et al. 2008; Kaplan 1995; Lidwell et al. 2010; Ulrich et al. 1991). Further, people tend to be more interested in natural scenes (Lidwell et al. 2010; Gao, Barbieri and Valdivia 2014), implying a greater likelihood to pay attention to images reflecting a natural environment.

Similarly, attention may be altered during emotional arousal, allowing people to adapt psychologically, physiologically, and behaviorally to the environment. In such a state attention mechanisms are automatically recruited (Lane et al. 1999). Emotion heightens sensitivity to visual cues (Lane et al. 1999) and as arousal increases, so does attention capacity (Coull 1998; Lang 1990). Thus, viewers may attend to the more highly arousing images more closely than to the less arousing images. Indeed, practitioners evoke positive emotional arousal strategically to engage consumers’ attention and retain their interest while viewing an ad. The effectiveness of such a strategy is evidenced in an eye-tracking study in which arousal states of surprise and joy successfully concentrated viewer attention (Teixeira, Wedel, and Pieters 2012), and in another study in which arousal scenes were associated with more viewing time and the preferential targets in viewing (LaBar et al 2000). However, no prior research has examined the effect of arousal on attention to tourism images.
Limited research has used eye-tracking to compare Asian and Western respondents. Chinese may be more inclined than Westerners to take a more holistic view of images, as Chinese attend to both foreground and background objects, whereas Americans focus more on foreground objects (Chua et al. 2005). Similarly, East Asians seem to possess a holistic cognitive style that processes a scene more globally than Westerners, who as analytical thinkers tend to detach objects from their wider context (Dong and Lee 2008; Nisbett 2003). Chinese are more likely to follow a circular pattern as a strategy for perceiving the webpage holistically, whereas Americans follow a sequential pattern, often starting from the center and moving to the periphery of the page (Dong and Lee 2008). Although these studies were not conducted in a tourism context, they provide sufficient basis to assume that ethnicity affects preferences and attention patterns in the aspects of marketing people attend to.

From the preceding review of literature, we conclude that attention paid to images is likely to differ on the basis of image characteristics (i.e., environment and level of arousal an image presents) and ethnicity (i.e., Australian vs. Chinese). More specifically, we propose:

*Hypothesis 1:* Natural images attract a higher number of fixations than built images.

*Hypothesis 2:* Higher arousal images attract a higher number of fixations than low arousal images.

*Hypothesis 3:* Natural images attract a longer duration of fixations than built images.

*Hypothesis 4:* High arousal images attract a longer duration of fixations than low arousal images.

*Hypothesis 5:* Attention and scanning patterns differ between Australians and Chinese, such that (a) the number of fixations differs between Australians and Chinese, (b)
the duration of fixation differs between Australians and Chinese, and (c) the
number of saccades differs between Australians and Chinese.

Research has considered recall and likability to be strong indicators of advertising
effectiveness (e.g., Mehta and Purvis 2006; Walker and Dubitsky 1994). Recall measures the
memorability of images but is criticized for favoring rational commercials over the more
emotional ones that characterize tourism promotions. Further, recall does not show the direction
of emotion (i.e., positive or negative). Likability on the other hand is strongly linked to emotions
and indicates travelers’ preference (Kastenholz and Young 2003). Both recall and likability are
measured in a self-report format incapable of revealing insights into how images are processed.
This study thus also seeks to answer the question: Do attention indicators, such as an increase in
fixation count and duration, correlate with self-reported liking and recall of images?

Method

This research involved a multi-measurement study using eye movements and self-reports. Before
commencing the main eye-tracking study, we subjected the images to be used to pre-testing so
that a total of 16 images could be presented as part of the final eye-tracking study.

Pre-testing Activities and Images

The first step entailed generating tourism activities that could be classified into one of the four
quadrants of our model. Descriptions for 60 tourism activities, generated from the literature (e.g.,
Pizam and Fleischer 2005), Tourism Australia’s annual International and Domestic Visitor
Surveys, and the travel magazine, Australian Traveller, were used to identify strength of arousal
and environment classification. A convenience sample of ten “tourists” and five tourism experts
rated each activity on five-point scales for arousal (from low to high) and environment (from
natural to built). Analysis of the data reduced the number of activities to be represented by photographic images to 20. Images were sourced from Tourism Australia’s collection of photographic images that are used by tourism businesses across Australia in their marketing activities. Using an online survey tool (Qualtrics), the researchers then randomly displayed selected images to a sample of 65, drawn from a marketing list company database, to determine the images that best represented the four quadrants.

Two questions adapted from Kaplan et al. (1972) and Mano and Oliver (1993) were asked about each image: The environment for engaging in an activity like the one shown in the image would be classified as: (1) Totally natural environment to (5) Totally built; and Overall, this activity would be classified as one that is: (1) Very low arousal to (5) Very high arousal, with a definition for arousal being stimulating and exciting. On the basis of its mean ratings for environment and arousal, each image was plotted within one of the four quadrants. In a second online survey, a sample of 331 participants, obtained with the assistance of the same marketing list company, further validated the images as well as items measuring environment, arousal, and liking. The pre-testing identified a total of 16 images, or four images representative of each quadrant, that were subsequently used in the eye-tracking study.

Main Study

Participants. Thirty undergraduate students (mean age=26 years, 12 males and 18 females) from an Australian university participated in the study. Fifteen were Chinese who had arrived in Australia in the preceding three months, and 15 were Australian residents of Caucasian ethnicity. Participants had normal (or corrected to normal) vision. Small samples are common in
eye-tracking studies and are not intended to be representative (Miellet et al. 2010; Pan et al. 2004; Pan et al. 2013 Wedel and Pieters 2008).

**Apparatus.** A Tobii T120 Eye Tracker, integrated into a 17-inch monitor, was used for the eye-movement task. This device uses infrared light to create reflection patterns on the corneas of participants’ eyes. These patterns are collected by infrared sensors and processed using algorithms to build a unique structural profile of the physical characteristics of each participant’s eyes. The reflections from the cornea can then be used to calculate the position of the eyes. The Tobii T120 is considered accurate and collects information every 8.3 milliseconds (ms).

**Procedure.** Before engaging in the eye-tracking exercise, participants were informed they would be viewing images reflecting tourism activities/attributes of Australia, but no specific task was assigned (thus allowing free viewing). Images were presented randomly to reduce any order effects of images. Each image was displayed for 12 seconds. Following the eye-tracking task, participants evaluated the images through an online questionnaire in Qualtrics. They listed the images they recalled and described aspects of taking a holiday in Australia that would most interest them. Background demographics were obtained and then each image was once again presented, with questions asking participants to rate the likability of each image (e.g., I really like this particular photo of a day spa) and the likability of the activity portrayed in the image (e.g., I really like the idea of going to a day spa) on five-point Likert scales anchored by strongly disagree (1) and strongly agree (5).
**Manipulation Checks**

Image classification was confirmed in the main study phase. The mean value of the natural condition is significantly different from that of the built condition ($p < .01$). Similarly, there is a significant difference in the mean value between the low- and high-arousal conditions ($p < .01$). A scatterplot of participants’ ratings of environment and arousal showed four distinct clusters of images corresponding to the four proposed quadrants. Further information on stimulus testing and manipulation check results is available upon request from the corresponding author.

**Data Analysis Approach**

We first analyzed the eye-movement data for images on the basis of number (fixation count), time (fixation duration), and space (saccades). A fixation typically lasts around 200–500 ms (Wedel and Pieters 2007). As clinical visuo-cognitive research commonly uses a threshold of 200 ms (Manor and Gordon 2003; Nikolaev et al. 2013), we excluded fixations below 200 ms from the analysis. Fixation count refers to the number of fixations made on an image or an area of the image, with fixation counts and where they were made providing information on which elements of the image are visually important. The duration of a fixation indicates how long the viewer maintains higher-level focus on the same region to allow for extraction of visual details (Goh, Tan, and Park 2009). For most of the analyses of the fixations, we used the first six seconds of viewing (Lane et al. 1999). Experts argue that in tasks such as scene perception, people grasp the meaning of a scene within the first few fixations and then start to fill in detail (Duchowski 2002), making the activity in the first six seconds likely to be the most insightful. We also present saccade data for the full exposure to an image. Owing to images being displayed one after another (and the first fixation being a residual of the previous fixation on an image), we
removed the first .02 second for each image to control for the landing point of participants’ first fixation. Data were then explored through self-reports.

**Results**

Analysis began with eye-tracking measures of attention, including fixation count, fixation duration, and saccades. The mean fixation count was 10.55 (SD = 1.19) across the 16 images, and the average duration of total fixation time across the six-second period was 3.56 seconds (SD = .47). For each dependent variable of fixation count, fixation duration, and saccades, we conducted a 2 (arousal) x 2 (environment) x 2 (ethnicity) mixed-design ANOVA, with arousal and environment being within subjects and ethnicity between subjects. We then analyzed the self-reported liking of the images in the same manner and illustrate Chinese and Australian data on selected images.

Table 1 summarizes significant effects revealed in the mixed ANOVA. All effects are reported at $p < .05$. A formal test of normality was performed using the Shapiro-Wilk test for small samples. Saccades were not normally distributed. The assumption of normality was satisfied after taking log transformation. Fixation durations and counts were generally normal, with some within-subject non-normality in the Australian group related to the high-arousal and low-arousal natural conditions. A further inspection of histograms suggested that the two series were approximately normal and therefore normality was assumed. Distributions of liking scores were normal.

*Table 1*
**Fixation Count**

First, we investigated participants’ number of fixations in the first six seconds of viewing the images. Tests of within-subjects and between-subjects effects revealed the effect of arousal, environment, and ethnicity. As shown in Table 1, we found no main within-subjects effect for the environment on fixation count, indicating similar patterns for built and natural environment images, rejecting H1. We found a main within-subjects effect of arousal on the number of fixations, supporting H2. A higher number of fixations were recorded for high-arousal images ($M = 10.78$) than for low-arousal images ($M = 10.32$). We found a main between-subjects effect for ethnicity on the number of fixations, supporting H5a. Australians had a higher mean fixation count than Chinese ($M_{\text{Australian}} = 11.45$ vs. $M_{\text{Chinese}} = 9.65$). We found no interaction effect for ethnicity and environment, but did find a significant interaction between arousal and ethnicity. Following this significant interaction effect, we performed simple-effects tests, which suggested that while Australians had virtually the same number of fixations for both low- and high-arousal images ($M_{\text{low arousal}} = 11.42$ vs. $M_{\text{high arousal}} = 11.47$), Chinese had a higher number of fixations for high-arousal images ($M_{\text{low arousal}} = 9.21$ vs. $M_{\text{high arousal}} = 10.09, p < .01$).

The three-way interaction between arousal, environment, and ethnicity was significant, indicating that the two-way interaction described for ethnicity and arousal varies with environment (see Figure 2). The pattern of the ethnicity by arousal interaction is similar for natural images (Australians: $M_{\text{low arousal}} = 11.43$ vs. $M_{\text{high arousal}} = 11.07$; Chinese: $M_{\text{low arousal}} = 8.93$ vs. $M_{\text{high arousal}} = 10.33, p < .01$). However, apparently no difference exists between Australian and Chinese participants irrespective of arousal under the built condition (Australians: $M_{\text{low arousal}} = 11.40$ vs. $M_{\text{high arousal}} = 11.88$; Chinese: $M_{\text{low arousal}} = 9.50$ vs. $M_{\text{high arousal}} = 9.85$).
Duration of Fixation

With respect to participants’ duration of fixations in the first six seconds of viewing, we found a main effect of arousal on the total duration of fixations. A longer fixation was recorded for high-arousal images \((M = 3.65\) seconds\) than for low-arousal images \((M = 3.47\) seconds\). A main effect occurred for the environment on fixation duration, indicating longer duration patterns for natural \((M = 3.65\) seconds\) versus built \((M = 3.47\) seconds\) environment images. These results support H3 and H4. H5b is also supported by a main effect for ethnicity on fixation duration. Australians had a longer mean fixation than did Chinese \((M = \text{Australian} 3.93\) seconds vs. Chinese 3.19 seconds\). We found no interaction effect for ethnicity and environment or between arousal and ethnicity for fixation duration, but did find an interaction effect between arousal and environment on fixation duration. For the arousal by environment interaction, duration of fixations was similar for low- and high-arousal images \((M = \text{low arousal} 3.46\) seconds vs. high arousal 3.49 seconds\) in the built image condition. However, the pattern varied in the natural image condition, with a higher duration of fixations recorded for high-arousal images \((M = \text{low arousal} 3.48\) seconds vs. high arousal 3.81 seconds, \(p < .01\)). Finally, the three-way interaction between arousal, environment, and ethnicity was significant, indicating that the two-way interaction described for arousal and environment varies with ethnicity (see Figure 3). Essentially, the pattern of the interaction holds for the Chinese group \((\text{Natural: } M = \text{low arousal} 2.96\) seconds vs. high arousal 3.48 seconds, \(p < .01\); Built: \(M = \text{low arousal} 3.17\) seconds vs. high arousal 3.16 seconds\) but not for the Australian group \((\text{Natural: } M = \text{low arousal} 4.01\) seconds vs. high arousal 4.15 seconds; Built: \(M = \text{low arousal} 3.74\) seconds vs. high arousal 3.82 seconds\).
As the saccades did not satisfy the normality assumption, we ran the test using log-transformed data and compared the results to results from the original saccade series. The two sets of results exhibited consistency in terms of significant effects. Given the difficulty of interpreting results produced using transformed data, this section presents results produced with the original series. We found no significant main within-subject effects for environment or arousal, but did find a between-subjects main effect for ethnicity. As Chinese had more saccades than Australians ($M_{\text{Chinese}} = 53.16$ vs. $M_{\text{Australian}} = 42.25$), H5c is supported. One interaction effect was evident: arousal by environment. As Figure 4 shows, this interaction reveals that in the low-arousal condition the natural versus built images show minimal difference between saccades ($M_{\text{natural}} = 48.75$ vs. $M_{\text{built}} = 47.21$), whereas the difference is significant in the high-arousal condition ($M_{\text{natural}} = 45.96$ vs. $M_{\text{built}} = 48.90$, $p < .05$).

### Australian versus Chinese Illustrative Patterns

To illustrate the different eye-movement patterns, we present a gaze plot for each group using a random sample of three participants (using the small sample reduced clutter). A gaze plot shows how participants’ eyes moved between spatial locations within an image and where participants directed their attention when viewing the images. As results for the rainforest example (low arousal and natural environment) make evident, Chinese had lower fixation counts and fixation durations (Figure 5) than Australians (Figure 6). Examination of the images shows that the number of direct fixations is lower for the Chinese, but the statistics for saccades suggest the
Chinese did more scanning without fixing as often as the Australians. Additionally, Australians’ eye movements covered a greater space within the image than Chinese.

Figures 5 and 6

Self-reported Liking and Recall

Respondents also self-rated how much they liked each of the 16 images on five-point Likert scales, yielding a mean liking score across the 16 images of 3.85 (SD = .46). A mixed within- and between-subjects ANOVA found that the main effect of arousal is not significant, but the main effect for environment was substantial. Natural images were better liked than built images ($M_{\text{natural}} = 4.03$ vs. $M_{\text{built}} = 3.70$). We found an interaction effect between arousal and environment on liking scores. Participants’ liking scores were similar for natural and built images in the high-arousal condition ($M_{\text{natural}} = 3.96$ vs. $M_{\text{built}} = 3.92$), but varied in the low-arousal condition, with a higher liking score for the natural images ($M_{\text{natural}} = 4.10$ vs. $M_{\text{built}} = 3.48$, $p < .01$). The effect of ethnicity is not significant. However, we found an interaction effect between ethnicity and environment. Australian participants had a similar level of liking for natural and built images ($M_{\text{natural}} = 3.89$ vs. $M_{\text{built}} = 3.73$), whereas Chinese scored higher in the natural condition ($M_{\text{natural}} = 4.16$ vs. $M_{\text{built}} = 3.66$, $p < .01$). Finally, the three-way interaction between arousal, environment, and ethnicity was not significant. Figure 7 presents the two-way interaction effects.

Figure 7

Association between Eye-tracking and Self-report Data

A correlation analysis shows that liking is not associated with mean fixation duration ($r = .019$, $p > .05$), mean fixation counts ($r = -.026$, $p > .05$), or total number of saccades ($r = -0.008$, $p > .05$). However, a high correlation ($r = .62$, $p < .01$) exists between liking of the image and liking of the
activity portrayed in the image. Table 2 summarizes mean fixation count, duration, saccades, liking, and recall by ethnicity and four quadrants.

Table 2

Scores for recall were calculated from the number of times an image appeared among participants’ top three most recalled images. High-arousal natural images were the most recalled image type and also the type that participants’ eyes dwelled on the longest, although they were not the most liked images. In contrast, low-arousal images in the built condition were the least liked, recalled, and attended to for Australian participants. They were also the least liked and recalled images for Chinese participants. However, these images did attract medium-level fixation durations. Further examination of respondents’ description of the aspect of holidaying in Australia that was most interesting to them revealed that eight of the 15 Chinese participants referred to the built environment, such as city and shopping, compared to only one Australian participant. This finding may explain the higher level of attention among Chinese to this image type.

Importantly, even when the two participant groups had similar liking scores regarding an image, areas of interest within the image based on which evaluation was made may differ. For example, Figures 8 and 9 show that the two groups paid attention to different regions, and that Australians had a much higher interest in the logo on the person’s chest (duration: $M_{\text{Australian}} = 0.87$ vs. $M_{\text{Chinese}} = 0.07$; count: $M_{\text{Australian}} = 0.5$ vs. $M_{\text{Chinese}} = 0.04$).

Figures 8 and 9


**Discussion and Conclusion**

In this study, eye-tracking and self-report data collected from Chinese participants were compared to data obtained from Australian participants in terms of responses to tourism photographic images.

The study yielded four main findings and contributions. First, attention to the photographic images varied, depending on whether the subject matter was a natural or built environment and a low or high level of arousal, confirming H2, H3, and H4 that natural and high-arousal images command greater attention. Second, supporting H5 regarding ethnicity’s effect, Australians and Chinese differed in terms of how attentive they were to tourism image stimuli and what elements of the images they attended to. Third, interactive effects demonstrate the complexity of factors that influence the way images are attended to. Fourth, attention to an image (higher fixation count or duration) is not correlated with liking of the image.

Consistent with prior research findings that attention capacity increases with arousal (Coull 1998; LaBar et al. 2000), our results show that high-arousal images received more fixation counts than low-arousal images, but we found no main effect for the natural/built environment. Fixation count can be an indicator of areas of importance or difficulty interpreting a task, although importance or difficulty can be partly determined by the task, such as a search task (Rakoczi et al. 2013). Our research did not specify a search or problem-solving task and was considered to be free viewing. More fixation counts in the high-arousal images suggest that these images contained more points of interest or things to focus on. In contrast, the low-arousal images were easier to comprehend and allowed more opportunity to gaze without fixating. Interestingly, the fixation count for arousal level of images was moderated by ethnicity. Chinese
participants had a higher fixation count in the high-arousal condition, suggesting that they saw more of importance in a high-arousal image. Furthermore, a three-way interaction effect demonstrated that this effect occurred predominantly in natural rather than built images.

The saccadic patterns also varied as a result of an interaction between arousal level and the environment, which occurred irrespective of ethnicity. More saccadic activity appears to be associated with the built environment when the image depicts high arousal. A reverse pattern occurs for the natural condition, with saccadic amplitude being greater in the low-arousal condition. In general, this finding suggests more searching in the built/high-arousal condition or the natural/low-arousal condition.

These findings are consistent with suggestions in the literature that nature and higher arousal can more effortlessly engage attention (Coull 1998; Kaplan 1995; Lang 1990; Lidwell et al. 2010; Ulrich et al. 1991), as more searching occurred when one of these conditions was not satisfied. Further investigation is required to advance the understanding of the underlying mechanisms for this finding. The saccadic amplitude also differed by ethnicity, which we discuss further in the section on Australian and Chinese differences.

Australians and Chinese differed overall in the way they evaluated the images. Australians’ higher fixation durations and fixation counts may be explained by their familiarity with the environment and the activities portrayed in the images. Research has found familiarity to be an important factor influencing attention (Karacan et al. 2010). Possibly the Australians were able to process information conveyed in the more prominent objects within an image more quickly than were the Chinese, which allowed the Australians to explore the less prominent but possibly engaging elements of the images. In contrast, Chinese participants’ gaze plots reflect
slower processing and more scanning with fewer fixations, possibly owing to unfamiliarity. This result is confirmed by the saccade data, which support the searching behavior of the Chinese.

Research proposing ambient and focal classification of fixations offers other general insights into the differences between Australian and Chinese patterns of fixation duration times and saccades (Velichkovsky et al. 2005). This distinction refers to the difference between a given fixation duration and the following saccade amplitude. A pattern of long fixation and short saccade may indicate a focal fixation (greater inspection), while a pattern of short fixation and long saccade indicates ambient fixation (greater exploration). Focal fixation is more cognitive, involving more inspection, whereas ambient fixation is more visual, involving exploration processing (Biele, Kopacz, and Krejtz 2013; Velichkovsky et al. 2005). Further, in processing information, individuals from Eastern countries focus more on context and those from the West focus more on objects, without particular attention to the context (Nisbett 2003; Dong and Lee 2008). This difference may help explain some of our ethnicity results, as the Australians focused more on objects whereas the Chinese did more scanning.

In conclusion, the results of our study contribute to the body of literature on tourism advertising’s effectiveness and demonstrate that the processing of visual tourism stimuli is complex and market segments may respond to photographic visual stimuli differently. Perhaps our most important finding is that while Australians fixate more often and for longer periods than the Chinese, this response varies with the image condition. The Chinese had low fixation counts and fixation duration periods, especially when viewing low-arousal natural images (e.g., rainforest walk, sunbathing on the beach), indicating that these images are perhaps less eye-catching to potential travelers from China. The greater number of saccades of Chinese suggests
different ways of processing tourism images, and the greater space Australians explored implies a need to consider the spatial distribution of elements in designing tourism visual stimuli.

**Practical Implications**

As consumers are exposed to an ever-growing amount of visual marketing, the ability to attract attention becomes increasingly important. Tourism marketers generally recognize the need to translate textual messages to the language and context of the market, but tend to ignore that markets may process and decode pictorial images differently.

This study provides insights regarding image-viewing behaviors that could assist in developing more effective stimuli and demonstrates the importance of tailoring images to specific target groups to solicit attention. Some images seem to be more effective in this effort. For instance, images portraying low-arousal activities in a natural environment received the least attention from the Chinese participants, while high-arousal images in a natural environment captured the most attention. As high-arousal images were the most recalled irrespective of ethnicity, further investigation might examine the relative effectiveness of image types in tourism promotion.

Notably, merely using images that consumers like may be inadequate to capture market attention. In contradiction to the common belief that participants look longer at stimuli they like and find motivating (e.g., Balcetis and Dunning 2006), our results reveal a lack of association between liking and attention. Potentially, arousal images may contain inherent confounds that make people respond in a unique way (Rupp and Wallen 2007). This finding highlights the need for future investigation of images on the basis of context and image types.
This study also suggests that a tourist who does not like an image may still engage with the image and extract meaning from it. Eye-tracking offers a window to attention and viewing patterns and is able to show what captures people’s interest, although not necessarily what they like. However, as eye-tracking reveals insights into people’s perception and cognition, efforts to evaluate marketing stimuli need to incorporate such measures.

Given visual attention’s influence on decisions such as brand choice (Atalay et al. 2012; Krajbich et al. 2010), attention-attracting marketing visuals can potentially increase visitations to a destination. We propose several guidelines for selecting/designing photographic images and other visual stimuli for tourism marketing:

- Use natural over built scenes;
- Use high-arousal over low-arousal scenes, especially for the Chinese market;
- If a destination is predominantly built, incorporate natural content (e.g., water and vegetation) in images portraying the destination;
- When promoting to the Australian market, key messages may be spread to the peripheral regions in stimuli, whereas for the Chinese, place these messages more toward the center;
- Select images according to the primary goal, for instance, getting a message across (attention), improving memorability (recall), or enhancing destination image (liking); and
- Research the audience’s attention patterns and viewing behaviors.

These implications may extend beyond marketing to the actual visual environment that travelers experience. For instance, during an on-site experience at the destination, travelers’
attention might be drawn to natural and high-arousal attractions, making these good platforms for promotional messages.

**Methodological Reflections for Tourism Research**

Methodologically, our approach differs from the survey- and interview-based approaches of most prior tourism marketing research. Verbal or written responses cannot adequately measure mental process such as scene perception, because individuals are not able to tell how they think. Even a measurement such as a scale measures only participants’ memory of the mental process rather than the actual process (Burns, Biswas, and Babin 1993). We have demonstrated that eye-tracking can provide valuable insights into consumers’ visual behaviors and offer an opportunity for data triangulation between physiological and self-report data. Unlike global measures such as liking, which focus on reaction to a holistic stimulus, eye-tracking enables researchers to examine reactions to isolated parts of the stimulus. We were able to identify images participants most attended to and where within an image they attended. Eye-tracking can be used in a variety of areas related to tourism, such as websites and assessments of the effectiveness of print materials such as menus, posters, brochures, and flyers.

The novel use of the eye-tracking methodology is this study’s strength but also its limitation. Although our study uniquely demonstrates that differences exist in viewing tourism marketing stimuli related to environment, arousal, and ethnicity, the difficulty of interpreting eye-tracking data allows only limited insight into the sources of these differences. The correlation between the two sets of liking scores and the lack of association between liking and attention leads us to speculate that liking of an image is determined by participants’ attitude
toward the activity portrayed in the image rather than by how much visual extraction occurred during the viewing.

However, Chinese and Australians are possibly not seeing the same thing when looking at the same image. As a result, similar liking scores may emerge from participants’ interest in different regions within an image. Whether our findings are unique to the images under investigation or reflect generalized group differences in viewing patterns also remains unclear, as inferring meaning from eye-tracking findings is inherently speculative (Rupp and Wallen 2007). To successfully deploy eye-tracking methodology, future studies must incorporate traditional interview or survey methods to collect detailed information on, for example, underlying motivations and interests that would help with the interpretation of eye-tracking findings.

Eye-tracking methodology also presents resource and logistic challenges. Excluding participants’ time and cost in getting to the research lab, an eye-tracking data collection session with a single participant typically lasts 30 to 60 minutes (in our case, 30 to 45 minutes), resulting in difficulty recruiting participants. Furthermore, as the eye-tracking device records participants’ scene-by-scene real-time eye movement in a large number of measures, the sheer quantity of data makes analysis time-consuming. Using images strictly manipulated for hypothesis testing would be more manageable, but researchers must balance between highly controlled images and images that are highly naturalistic. In addition, no standardized attention metrics exist to provide a benchmark for the results for either academic research or marketing practice.

While our study does not explain all of the issues relating to differences in response to pictorial imagery stimuli, it does suggest that different types of attention are engaged for different types of tourism images and for different groups of potential tourists. To our
knowledge, this study is the first to show an ethnic difference in processing and evaluating tourism images. The study points to a new direction in research regarding visuals in tourism promotion, and provides a basis for further research to understand the underlying mechanisms of preference for and recall of tourism visuals. The strategy adopted in this study demonstrates a workable approach, offering promise of a better understanding of differentiated consumer response to marketing materials as well as the usability of eye-tracking to evaluate the effectiveness of promotional materials in tourism.

**Limitations and Future Research**

A major limitation of this study is its low statistical power as a result of the small sample used, reducing the chance of detecting a true effect, and other effects may have gone undetected in this study. Nevertheless, the significant differences revealed with this small sample have a high reproducibility with larger samples. Other key limitations of the present research include the use of photographic images rather than real tourism scenes and the use of photos only with no text, and the absence of a specific task (free viewing).

Future studies could use different stimuli and also combine more real search activities used by tourists. Using images in conjunction with text could be insightful, and the interesting findings in the ethnic variation of eye-tracking patterns lead to the possibility of different scene processing (focal vs. ambient). Recent research into ambient and focal attention suggests that a more positive mood may lead to focal attention and a neutral mood to more ambient attention (Biele et al. 2013). Future research could include measures of mood. Additionally, the variance in viewing patterns and attention identified by this study may result from multiple psychological and biological factors such as age, personality, familiarity with the context, and experience with
visuals, another topic worthy of further investigation. Finally, future research might use the eye-tracking method to investigate advertisement effectiveness and travelers’ decision making.
References


**Table 1. Summary of Test Results**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Fixation Count</th>
<th>Fixation Duration</th>
<th>Saccades</th>
<th>Liking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F(1, 28)$</td>
<td>$\eta_p^2$</td>
<td>$F(1, 28)$</td>
<td>$\eta_p^2$</td>
</tr>
<tr>
<td>Environment$^a$</td>
<td>ns</td>
<td>Ns</td>
<td>8.53**</td>
<td>.23</td>
</tr>
<tr>
<td>Arousal$^a$</td>
<td>6.09*</td>
<td>.18</td>
<td>12.29**</td>
<td>.30</td>
</tr>
<tr>
<td>Ethnicity$^b$</td>
<td>8.76**</td>
<td>.24</td>
<td>9.77**</td>
<td>.26</td>
</tr>
<tr>
<td>Environment x Arousal</td>
<td>ns</td>
<td>Ns</td>
<td>7.47**</td>
<td>.21</td>
</tr>
<tr>
<td>Environment x Ethnicity</td>
<td>ns</td>
<td>Ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Arousal x Ethnicity</td>
<td>4.67*</td>
<td>.14</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Environment x Arousal x Ethnicity</td>
<td>4.66*</td>
<td>.14</td>
<td>4.85*</td>
<td>.15</td>
</tr>
</tbody>
</table>

Note: *$=p<.05$; **$=p<.01$. $^a$ denotes a within-subject factor and $^b$ denotes a between subject factor.
Table 2. Fixation Count, Duration, Saccades, Liking and Recall by Ethnicity and Four Quadrants

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Stimulus</th>
<th>LAN</th>
<th>HAN</th>
<th>LAB</th>
<th>HAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian</td>
<td>Fixation count</td>
<td>11.43</td>
<td>11.07</td>
<td>11.40</td>
<td>11.88</td>
</tr>
<tr>
<td></td>
<td>Fixation duration</td>
<td>4.01</td>
<td>4.15</td>
<td>3.74</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td>Saccades</td>
<td>42.57</td>
<td>39.80</td>
<td>43.13</td>
<td>43.5</td>
</tr>
<tr>
<td></td>
<td>Liking</td>
<td>3.97</td>
<td>3.85</td>
<td>3.48</td>
<td>3.85</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>3.5</td>
<td>3.5</td>
<td>0.5</td>
<td>3</td>
</tr>
<tr>
<td>Chinese</td>
<td>Fixation count</td>
<td>8.93</td>
<td>10.33</td>
<td>9.50</td>
<td>9.85</td>
</tr>
<tr>
<td></td>
<td>Fixation duration</td>
<td>2.96</td>
<td>3.48</td>
<td>3.17</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>Saccades</td>
<td>54.93</td>
<td>52.12</td>
<td>51.28</td>
<td>54.31</td>
</tr>
<tr>
<td></td>
<td>Liking</td>
<td>4.20</td>
<td>4.12</td>
<td>3.33</td>
<td>3.98</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>1</td>
<td>3.5</td>
<td>1</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Figure 1. Tourism activities by arousal/environment quadrants
Figure 2. Mean fixation count – three-way interaction
Figure 3. Mean fixation duration – three-way interaction
Figure 4. Saccades – two-way interaction
Figure 5. Gaze plots for Chinese participants (n=3); image by Tourism Australia

Figure 6. Gaze plots for Australian participants (n=3); image by Tourism Australia
Figure 7. Liking – two-way interactions
Figure 8. Heap map of Australian participants (n=15); image by Tourism Australia

Figure 9. Heap map of Chinese participants (n=15); image by Tourism Australia