



Understanding Consumers' Visual Attention in Mobile Advertisements: An Ambulatory Eye-Tracking Study with Machine Learning Techniques

Wen Xie^a , Mi Hyun Lee^b , Ming Chen^c , and Zhu Han^a 

^aUniversity of Houston, Houston, Texas, USA; ^bNorthwestern University, Evanston, Illinois, USA; ^cUniversity of North Carolina, Charlotte, Charlotte, North Carolina, USA



ABSTRACT

As mobile devices have become a necessity in our daily lives, mobile advertising is also prevalent. Accordingly, it is critical for practitioners to understand how consumers visually attend to mobile advertisements. One popular way of doing so is via eye-tracking methodology. However, scant eye-tracking research exists in mobile settings due to technical challenges, e.g., cumbersome data annotation. To tackle these challenges, the authors propose an object-detection machine learning (ML) algorithm—You Only Look Once (YOLO) v3—to analyze eye-tracking videos automatically. Moreover, we extend the original YOLO v3 model by developing a novel algorithm to optimize the analysis of eye-tracking data collected from mobile devices. Through a lab experiment, we investigate how two types of ad elements (i.e., textual vs. pictorial) and shopping devices (i.e., mobile vs. PC) affect consumers' visual attention. Our findings suggest that (1) textual ad elements receive more attention than pictorial ones, and such differences are more pronounced in ads on mobile devices than those on PCs; and (2) mobile ads receive less attention than PC ads. Our findings provide managerial insights into developing effective digital advertising strategies to improve consumers' visual attention in online and mobile advertisements.

Mobile devices have become prevalent in our daily lives, and thus have developed into an essential digital advertising channel. The Pew Research Center's (2021) survey shows that 85% of Americans own smartphones; such high usage is consistent across different age groups (95% of those aged 18 to 49 years, 83% aged 50 to 64, and 61% over 65). Accordingly, mobile advertising has grown. A report from Statista (Chevalier 2022) indicates that, as of the second quarter of 2022, retail website orders using smartphones in the United States reached 60%, much higher than those using desktops (38%) or tablets (2%). Additionally, a recent report from the Interactive Advertising Bureau (IAB 2023) revealed that the mobile channel accounted for 73.5% (equivalent to \$154.1 billion) of internet advertising revenues in 2022. Therefore, it is critical to understand how

consumers attend to mobile advertisements to develop more effective digital advertising strategies.

Extant literature (Pieters and Wedel 2004; Chun and Wolfe 2005) has shown that consumers paying attention to visual stimuli while shopping online is influenced by two types of factors: (1) bottom-up factors, which reside in the stimulus; and (2) top-down factors, which relate to the attentional process. The salience of visual stimuli affects their received attention because salient stimuli are more noticeable and stand out more easily from other nearby stimuli (Van der Lans, Pieters, and Wedel 2008). Perceptual features such as the layout and size of stimuli determine their salience (Janiszewski 1998; Pieters and Wedel 2004). The informativeness of visual stimuli impacts how long consumers pay attention to the stimuli. Specifically, informativeness influences the speed at

CONTACT Mi Hyun Lee  mihyun.lee@northwestern.edu  Medill School of Journalism, Media, Integrated Marketing Communications, Northwestern University, 1845 Sheridan Road, Evanston, IL 60208, USA.

Wen Xie (BEing and BEcon, University of Electronic Science and Technology of China) is a doctoral candidate, Department of Electrical and Computer Engineering, Cullen College of Engineering, University of Houston.

Mi Hyun Lee (PhD, Arizona State University; PhD, Virginia Tech) is an assistant professor, Integrated Marketing Communications, Medill School of Journalism, Media, Integrated Marketing Communications, Northwestern University.

Ming Chen (PhD, University of Houston) is an assistant professor, Department of Marketing, Belk College of Business, University of North Carolina, Charlotte.

Zhu Han (PhD, University of Maryland) is the John and Rebecca Moores Professor, Department of Electrical and Computer Engineering, Cullen College of Engineering, University of Houston.

Copyright © 2023, American Academy of Advertising

which consumers acquire information that is contingent with their goals, which involves slow, serial processes (Pieters and Wedel 2007).

Applying the theory of attention described above, the current study investigates how online advertising elements (e.g., product descriptions and images) and shopping devices (e.g., mobile and PC) affect consumers' visual attention. When consumers shop online, different types of ad elements (i.e., textual and pictorial) may have unequal salience and informativeness such that they receive different levels of visual attention (Rayner et al. 2001; Wooley et al. 2022). Additionally, the advertisement sizes on mobile device screens are smaller than those on PC screens due to the limited display space (Ghose, Goldfarb, and Han 2013). These different ad sizes affect the ad elements' salience and informativeness, and thus their received attention (Peschel and Orquin 2013). Moreover, the same ad elements could have distinct salience and informativeness across PC and mobile ads due to their different sizes and layouts (Krider, Raghubir, and Krishna 2001). Therefore, it is practically and academically important to examine how different ad elements and shopping devices affect consumers' visual attention in online shopping environments.

To this end, we collected a novel data set using a portable (wearable) eye-tracking device (see Figure 1) in a lab experiment conducted in two settings: one with participants shopping online using a PC and the second with participants using a mobile device. We operationalized the visual attention that each ad element receives using consumers' eye-fixation count and fixation duration (Just and Carpenter 1980; Pieters and Wedel 2007; Wooley et al. 2022). With 132 participants, we analyzed a rich data set to demonstrate how two ad element types (i.e., textual [text, price, and rating] and pictorial [image]), two shopping

device types, (i.e., mobile and PC), and their interactions affect consumers' visual attention.

However, two major challenges remain in analyzing eye-tracking data, especially data collected using mobile devices. First, visual stimuli (i.e., ad elements) on mobile screens are not static: consumers may not always hold their mobile devices steady while using them, thereby complicating eye-tracking data analyses (Scott et al. 2019). Manually annotating these dynamic stimuli can be time-consuming or even infeasible (Meißner et al. 2019). Second, the small screen size of mobile devices results in relatively small visual stimuli. Accurately identifying small objects from eye-tracking videos is technically demanding, as it is difficult to distinguish small objects from the background (Tong, Wu, and Zhou 2020). Most existing eye-tracking research has focused on either static areas of interest (AOIs) on PCs (e.g., Maslowska et al. 2020) or screen-based (fixed) eye-tracking devices (e.g., Pfiffelmann, Dens, and Soulez 2020). Screen-based eye trackers impose restrictions on participants' head positions (Carter and Luke 2020), making them unsuitable for use with mobile devices; instead, portable eye-tracking devices (which do not impose such constraints) are required.

To overcome these challenges, the current study introduces an object-detection method—You Only Look Once (YOLO) v3 (Redmon and Farhadi 2018)—to automatically annotate eye-tracking data. Notably, YOLO v3 has shown superior performance in terms of accuracy and efficiency, as highlighted by Jayawardena and Jayarathna (2021). Furthermore, we extend the original YOLO v3 (one-step YOLO hereafter) and develop a two-step YOLO that detects small objects accurately in eye-tracking data collected from mobile devices. By leveraging these two YOLO approaches, we can obtain visual metrics

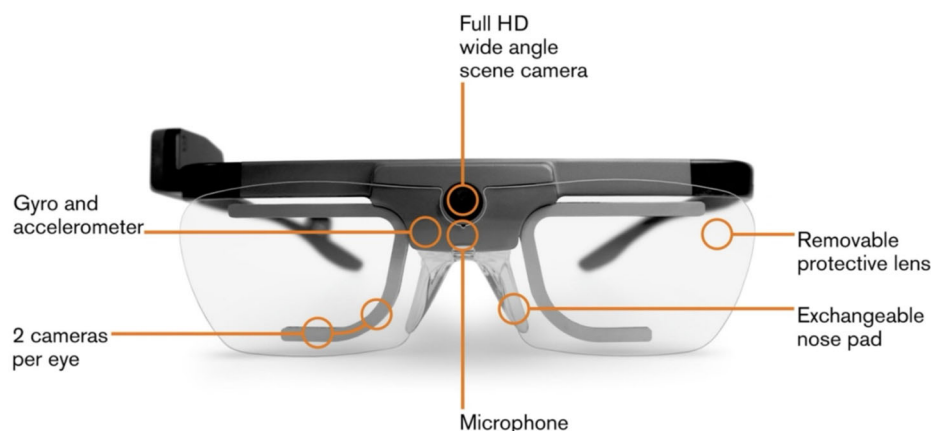


Figure 1. Ambulatory eye-tracking device (Tobii Pro Glass 2). (This figure is a screenshot from the website of the Tobii company.)

(e.g., eye-fixation count and duration) from the data collected using portable eye trackers from both PC and mobile devices, thereby enabling us to address our research objective.

Our contributions are fourfold. First, we offer new insights into the theory of attention (Pieters and Wedel 2007; Wooley et al. 2022), as we empirically examine how consumers' visual attention varies across the types of ad elements (i.e., textual vs. pictorial) and shopping devices (i.e., mobile vs. PC). Second, we develop a novel machine learning (ML) approach, the two-step YOLO, to automatically detect objects via eye-tracking data collected from mobile devices. Specifically, we propose a one- and two-step YOLO, each applicable for eye-tracking data collected from PCs and smartphones, respectively, enabling us to study consumers' visual attention across PC and mobile devices. Third, this study contributes to the consumer search domain (Zhu and Dukes 2017; Wang et al. 2019) by unveiling the distributional changes in consumers' visual attention as they progress from an earlier stage of their shopping journey (product searches) to a later one (making purchase decisions). Finally, our findings offer managerial implications for practitioners and online advertising platforms on how to reach audiences more effectively and develop cross-device strategies.

Literature Review

Theory of Attention: Determinants and Drivers of Attention

The research on visual attention traces back to James (1890), which described the attention process as "withdrawal from some things in order to deal effectively with others." James (1890) categorized attention control factors into two types: passive (i.e., bottom-up) and active (i.e., top-down). Bottom-up factors pertain to advertisement features that determine their perceptual salience (e.g., size and shape), while top-down factors reside in an individual's attentional processes (e.g., product involvement) (Pieters and Wedel 2004).

The salience of visual stimuli affects bottom-up attention (Yantis and Jonides 1990). A stimulus with salient features can receive attention automatically, even without conscious intent. Indeed, people usually shift their attention successively to stimuli with decreasing salience (Van der Lans, Pieters, and Wedel 2008). Bottom-up features, such as size and layout, play a crucial role in determining the perceptual salience of stimuli in an environment (Janiszewski 1998).

A stimulus that contrasts with its surroundings around perceptual features is visually salient and receives more attention. Moreover, a larger-sized stimulus is more salient than its smaller-sized counterpart. Hence, larger-sized stimuli receive more attention (Janiszewski 1998). For instance, larger-sized textual elements in print advertisements receive more attention (Pieters and Wedel 2004). A stimulus' layout can also affect its salience. For example, a stimulus in a square layout is more salient than the same one in a rectangular layout whose horizontal dimension is shorter than the vertical edge. This is because human beings perceive the horizontal dimension as the primary dimension when comparing squares with rectangles (Krider, Raghubir, and Krishna 2001). As a result, when consumers shop online, they tend to allocate their attention to salient ad elements with large sizes or in a square layout in a rapid, transient time course (Chun and Wolfe 2005).

A stimulus's informativeness affects top-down attention, which impacts how long consumers focus their attention on the stimulus to acquire information and make decisions (Pieters and Wedel 2007; Orquin and Loose 2013). This process involves slow, serial processes, resulting in a slow, sustained time course (Chun and Wolfe 2005). While more informative stimuli can receive more attention, a stimulus' informativeness depends on consumers' specific goals (Pieters and Wedel 2007). Specifically, these goals guide consumers' visual searching process and attention in terms of where they would look to find their desired information (Janiszewski 1998). As different ad elements convey different information, consumers allocate unequal attention to them under different goals. For instance, when the goal involves learning about a brand or making a purchase, consumers typically pay more attention to textual ad elements and less attention to pictorial ones (Rayner et al. 2001; Pieters and Wedel 2007; Pan, Zhang, and Law 2013). On the other hand, when the goal involves evaluating or comparing products, consumers tend to spend more time looking at pictorial ad elements than textual ones (Rayner, Miller, and Rotello 2008).

In this study, we focus on four ad elements (text, image, rating, and price) that capture consumers' visual attention in an online hotel shopping environment. We do so for two reasons. First, each ad element represents a specific attribute that carries distinct information. Second, most online ads contain a pictorial presentation of a brand or product (i.e., image) and textual description elements (i.e., text, price, and rating) (Wedel and Pieters 2006). Thus, we

refer to images as pictorial ad elements and the text, price, and rating as textual ad elements hereafter. Pictorial ad elements typically exhibit higher salience than textual ones, as they are more noticeable (Pieters and Wedel 2007). However, during online shopping, textual ad elements (such as the product name, price, and other descriptions) provide more product information, making them more informative than pictorial ones and drawing consumers' attention to them (Rayner et al. 2001; Pieters and Wedel 2007). Thereby, we develop the following hypothesis:

H1: Textual ad elements (i.e., text, price, and rating) receive more attention than pictorial ad elements (i.e., images).

PCs and mobile devices have been the dominant online shopping devices with their widespread usage. As a result, it is crucial to examine consumers' visual attention to advertisements across the two device types in online shopping environments. A notable difference between PC and mobile ads is their size, with PC ads being larger and mobile ads smaller due to their limited screen size (Ghose, Goldfarb, and Han 2013; Wang et al. 2019). Larger ads receive more attention than smaller ones (Lohse 1997), not only due to their increased salience (Peschel and Orquin 2013) but also because people require more attentional demand when viewing larger-sized ads (Pieters and Wedel 2004). Furthermore, ad elements (e.g., images) in mobile ads may appear cropped along the horizontal dimension due to the shorter width of mobile screens (note that this phenomenon is ubiquitous in the mobile apps of major travel agencies such as Booking.com, Expedia.com, and Hotels.com). This cropped appearance attenuates the overall informativeness and leads consumers to pay less attention to ads on mobile devices versus PCs. Hence, we hypothesize:

H2: Mobile ads receive less attention than PC ads.

We further investigate the interaction effects between ad elements and device types on visual attention: the same ad elements could have different levels of informativeness or salience between mobile devices and PCs due to the different screen sizes and ad element layouts. Even if ad elements have the same level of informativeness across PCs and mobile devices, their salience tends to differ significantly, especially for pictorial elements. This is because the bottom-up features in pictorial ad elements (i.e., image layouts) play a more critical role in decreasing their salience on mobile devices (compared to PCs), relative to the

features in textual elements. Pictorial ad elements have a square layout in PC ads but have a rectangular layout on mobile screens, making them less salient on mobile devices because they lose their salience due to the rectangular layout (Krider, Raghubir, and Krishna 2001). Thus, we expect shopping on mobile devices to strengthen the positive effect of textual ad elements on attention (i.e., receiving more attention than pictorial ones).

Furthermore, textual ad elements have the same content on both PC and mobile ads such that they have the same level of informativeness across devices. However, the pictorial ad elements displayed on mobile devices lose more informativeness than textual ones because they appear incomplete or distorted due to the small screen sizes (Wang et al. 2019), even with the same level of salience on both mobile and PC screens. Additionally, pictorial ad elements appear cropped along the horizontal direction to fit the rectangular layout on mobile ads, which may hide large parts of the images displayed on PCs with a square layout (for example, see Figures 4 and 5), causing them to be less informative on mobile devices than on PCs. Thus, we develop the following hypothesis and summarize all of the hypotheses in Figure 2.

H3: The effects of ad element types (i.e., textual vs. pictorial) and shopping device types (i.e., mobile vs. PC) on attention interact such that shopping on mobile devices strengthens the positive effect of textual ad elements (i.e., text, price, and rating) on attention (i.e., receiving more attention than pictorial ad elements [i.e., images]), but shopping on PCs attenuates it.

Eye-Tracking Studies in Advertising

Researchers in advertising, media, and marketing fields have conducted eye-tracking studies to investigate people's visual attention in different contexts, for example, personalized advertising (Pfiffelmann, Dens, and Soulez 2020; Segijn, Voorveld, and Vakeel 2021), media multitasking (Beuckels et al. 2021), online banners (Hernández-Méndez and Muñoz-Leiva 2015), online reviews (Maslowska et al. 2020), visual complexity (Pieters, Wedel, and Batra 2010), taboo advertising (Myers et al. 2020), print advertising (Simola, Kuisma, and Kaakinen 2020), video advertising (Wooley et al. 2022), influencer advertising (Pozharliev, Rossi, and Angelis 2022), ad avoidance (Schmidt and Maier 2022), online choice environments (Shi, Wedel, and Pieters 2013), and retail settings (Meißner et al. 2019; Pfeiffer et al. 2020; Chen

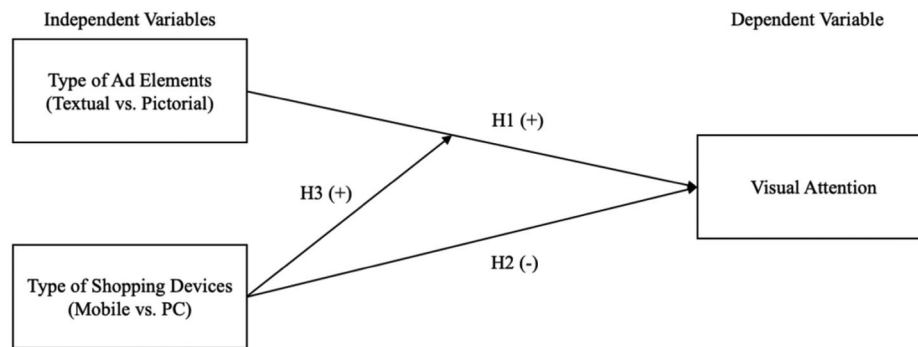


Figure 2. Conceptual framework of the developed hypotheses. The signs in parentheses next to each hypothesis indicate the effect of the first category on the dependent variable, compared to the second category. For example, hypothesis 1 (+) shows that textual ad elements receive more attention than pictorial ad elements.

et al. 2021). However, even among the most recent eye-tracking studies cited above, only a few have conducted experiments with mobile devices, each of which has limitations. Segijn, Voorveld, and Vakeel (2021) used tablets, which are less portable and less commonly used than smartphones. Pozharliev, Rossi, and Angelis (2022) employed a screen-based eye-tracking device and its accompanying mobile stand to adapt its use to mobile devices, which might constrain participants' head position (Carter and Luke 2020) and restrict the mobility of using mobile devices. Schmidt and Maier's (2022) study did not involve annotating small AOIs on mobile screens.

The scarcity of eye-tracking studies on mobile devices mainly stems from technical difficulties in annotating the eye-tracking data of users' visual focus on small mobile screens. Most existing eye-tracking studies rely on manual annotation in professional software (e.g., Shi, Wedel, and Pieters 2013; Pfeiffer et al. 2020; Chen et al. 2021; Pozharliev, Rossi, and Angelis 2022). However, manually annotating data collected from portable eye-tracking devices is time-inefficient (Meißner et al. 2019). When eye-tracking devices are portable, the AOIs' locations vary across the recorded video streams (Scott et al. 2019). This requires a more complex level of annotation: it is necessary to adjust an AOI's manually annotated location when it moves in each video frame. As noted in Chen et al.'s (2021) study, the manual annotation of a 30-minute eye-tracking video frame by frame "requires about 4 h" (p. 19), and in total, over "700 h" (p. 19). This is not economically feasible for most eye-tracking videos that last for hours. To tackle this challenge, automatic annotation is desirable, but reliable algorithms are not apparent in advertising or marketing research (Meißner et al. 2019).

To address this issue, the current study introduces YOLO (You Only Look Once) v3 (Redmon and

Farhadi 2018), an object-detection technique. By implementing YOLO v3 in their eye-tracking experiments, researchers can advance the eye-tracking data process toward an automated system, substantially saving time and labor costs.

Automatic Detection of AOIs in Eye-Tracking Studies

In eye-tracking studies, annotating consumers' visual focus is crucial in obtaining visual metrics (e.g., eye-fixation count). Manual annotation is feasible when a study involves a few static AOIs with fixed locations in the recorded eye-tracking videos, thus minimizing labor costs. However, annotating data collected from portable eye-tracking devices presents challenges, even for the auto-mapping tools in some expensive professional software such as Tobii Pro Lab. For example, it is cumbersome to annotate repeatedly recorded identical AOIs (Brône, Oben, and Goedemé 2011).

To combat this challenge, researchers have deployed machine learning (ML) techniques to automate AOI annotation in eye-tracking data. Some studies have recommended using traditional object recognition algorithms such as SIFT (scale-invariant feature transform) (e.g., Kurzhals et al. 2017). With the emergence of deep learning, more advanced object-detection methods have been developed to identify object locations in images or videos (Liu et al. 2020). For example, Hessels et al. (2018) and Jongerius et al. (2021) used OpenPose, an open-source object-detection library, to detect human faces in eye-tracking videos, showing comparable results to human annotators. Because OpenPose can only detect the human body, face, hands, and footprints (Cao et al. 2021), researchers have employed various advanced models to analyze dynamic AOIs in eye-tracking videos such as YOLO v2 (Callemein et al. 2018), ResNet and Mask R-CNN (Region-Based

Table 1. Literature on the object-detection methods used in analyzing eye-tracking data.

Study	Context	Eye-Tracking Device Type	Data	Types of Objects		Detection Method	Methodological Contribution
				Pictorial Object	Textual Object		
Kurzahls et al. (2017)	Print advertising	Portable	15 recordings	Image and product	Title, text, and price	K-means with manual verification	Four times faster than manual annotation with comparable accuracy
Hessels et al. (2018)	Social interaction	Screen-based	A public dataset	Human faces	N/A	OpenPose	OpenPose is as effective as semi-automatic approaches
Jongerius et al. (2021)	Consultation	Portable	Seven recordings	Human faces	N/A	OpenPose	High interrater agreement between human annotators and OpenPose
Callemein et al. (2018)	Face-to-face interaction	Portable	Three public datasets	Human hands and heads	N/A	YOLO v2 and OpenPose	Both models outperform traditional approaches
Barz and Sonntag (2021)	Video viewing	Screen-based	A public dataset	17 AOIs, e.g., car	N/A	ResNet and Mask R-CNN	The models perform well only for AOIs with distinct concepts and strong matches to the pre-trained objects
Jayawardena and Jayarathna (2021)	Video viewing	Portable	A public dataset	Four AOIs, e.g., bus	N/A	Faster R-CNN and YOLO v3	The models achieve similar accuracy, but YOLO v3 is much faster
This article	PC and mobile advertising	Portable	235 recordings	Image	Text, rating, and price	YOLO v3 and two-step YOLO	Accurately detect both pictorial and textual AOIs; the two-step YOLO can detect small-sized AOIs on smartphones and is 150 times faster than manual annotation

Convolutional Neural Network) (Barz and Sonntag 2021), and Faster R-CNN, and YOLO v3 (Jayawardena and Jayarathna 2021).

In advertising research, advertisements typically contain pictorial (e.g., product photos) and textual (e.g., brand name) elements. However, existing studies (e.g., Callemein et al. 2018; Jongerius et al. 2021) have focused on exploring the auto-detection of pictorial objects (e.g., human faces) in eye-tracking data. Detecting textual objects poses greater challenges, as they are less salient than pictorial objects, and their content can vary even within the same object type (e.g., the brand name differs from one brand to another) (Ye and Doermann 2015). In this study, our goal was to detect both pictorial (e.g., image) and textual (e.g., text, price, and rating) objects in online advertising. To achieve this, we selected the YOLO v3 model due to its ability to accurately detect objects in real time, outperforming other ML-based algorithms such as Faster R-CNN (Jayawardena and Jayarathna 2021).

Based on YOLO v3, we further proposed a two-step YOLO approach optimized for AOI detection on smartphones. AOI sizes are small on mobile devices (e.g., smartphones) and become much smaller in eye-tracking videos because, in natural settings, people view mobile screens at a distance. Successfully detecting small objects is challenging due to the lack of

appearance-related information required to distinguish them from the background or similar objects (Tong, Wu, and Zhou 2020). Our proposed two-step YOLO method overcomes these challenges, making it feasible to accurately detect small-sized AOIs on mobile devices. To the best of our knowledge, the current study is the first to apply ML techniques optimized for annotating eye-tracking data collected from mobile devices in advertising research. This application allows us to efficiently and accurately study how consumers' visual attention patterns across different ad elements vary between PC and mobile devices. In Table 1, we summarize the literature on the object-detection methods used in analyzing eye-tracking data and compare them to our study.

Method

To test our hypotheses, we obtained consumers' visual metrics (i.e., eye-fixation count and duration) from eye-tracking data through a lab experiment in three steps. Figure 3 illustrates our methodological framework. In Step 1, we collected the eye-tracking data during each participant's online shopping session. In Step 2, we detected the AOIs' locations using the proposed YOLO models. In Step 3, we matched the model-detected AOIs with participants' eye-fixation

coordinates and computed the attentional measures. This approach allowed us to obtain the eye-fixation count and duration for each AOI, enabling us to study consumers' visual attention across different ad elements and shopping devices. We introduce the details of each step in the following subsections.

Eye-Tracking Experiment

We conducted the lab experiment in two settings (i.e., PC and mobile devices) using a portable eye-tracking device—the Tobii Pro Glass 2 (see Figure 1), which researchers have commonly used in eye-tracking studies (Schmidt and Maier 2022). As we illustrate in Figure 1, the device had a scene video camera embedded in the glass frame, recording what appeared in the visual range while participants shopped online (e.g., advertisements, smartphones, and PCs). Further, the device had four sensors underneath the participants' eyes to capture their eye movements. We next provide details of the two experimental settings.

Setting 1: PC-Based Portable Eye-Tracking Experiment

We first collected a data set of consumers' visual attention to PC web ads. This setting took place at a research university in a major U.S. metropolitan city from August to November 2019. Fifty-three subjects (35 males and 18 females) participated after meeting the criteria of being 18 years old or older and not wearing prescription glasses. All participants had prior experience in booking hotels using PCs before the experiment. We selected hotel shopping in the experiment because the availability of many hotel options is beneficial for training our proposed YOLO models to detect different ad elements. We selected two destinations: (1) Austin, TX, as a representative business-trip destination because this city has various industries (e.g., automotive and e-commerce) and major high-tech corporations (e.g., Tesla and Amazon); and (2) Hawaii, as a representative leisure trip destination. We chose Booking.com as the shopping platform, as it is one of the most popular online travel agencies.

Before the experiment, participants viewed the following instructions: "Please sign on the consent form and fill out a survey about demographic questions. Please choose at least one of the two destinations (Hawaii and Austin, TX) to book a hotel using Booking.com, which is presented on a Dell laptop. During your online shopping journey, you can scroll with the mouse to move the web page up and down continuously and naturally to browse more hotel options. Once you make your final decision about the

hotel, please let the research assistant know. You can take as much time as you need." Next, a research assistant instructed them to wear the Tobii Pro Glass 2, followed by calibration. The Tobii Pro Glass Controller indicated whether the calibration was successful or not. After calibration, they proceeded to shop for hotels as they would normally do.

Out of the 53 participants, 28 booked hotels for both Austin, TX, and Hawaii, while 25 booked only one hotel. This yielded a total of 81 eye-tracking videos, with each representing a single shopping session. We excluded two recordings due to a false calibration, resulting in a total of 79 eye-tracking videos with an average length of 3.5 min. The eye-tracking videos included a recording of what appeared in the field of view of the Tobii Pro Glass 2 scene camera during the shopping session. To study consumers' visual attention to PC web ads, we focused on the following four AOIs of online advertisements: image (pictures of the hotel options), rating (consumer ratings or review scores), price (booking prices), and text (textual information, i.e., hotel names or addresses). For illustration purposes, Figure 4 displays an example of the four AOIs from a PC web advertisement.

Setting 2: Smartphone-Based Portable Eye-Tracking Experiment

We then gathered data on consumers' visual attention to smartphone ads, which took place from October to November 2021 in the same lab. Among a total of 79 participants who met the same screening criteria, 62 were male, and 17 were female. Eighty-three percent of the participants had prior experience in using smartphones to book hotels before participating in this experimental setting. Before the experiment began, participants received similar instructions with additional information: "Please choose at least one of the two destinations (Hawaii and Austin, TX) to book a hotel using the Booking.com app installed on a smartphone—iPhone 11. During your online shopping journey, you can use your fingers to move the search result page up and down continuously and naturally to browse more hotel options." Then, participants proceeded to start shopping for hotels using the smartphone after putting on and calibrating the eye-tracking device.

In total, we collected 154 eye-tracking videos, as 75 out of the recruited 79 participants chose to book hotels for both destinations. We excluded six recordings that did not pass the calibration, resulting in a final sample of 148 eye-tracking videos with an average length of 2.1 min. When the second experimental

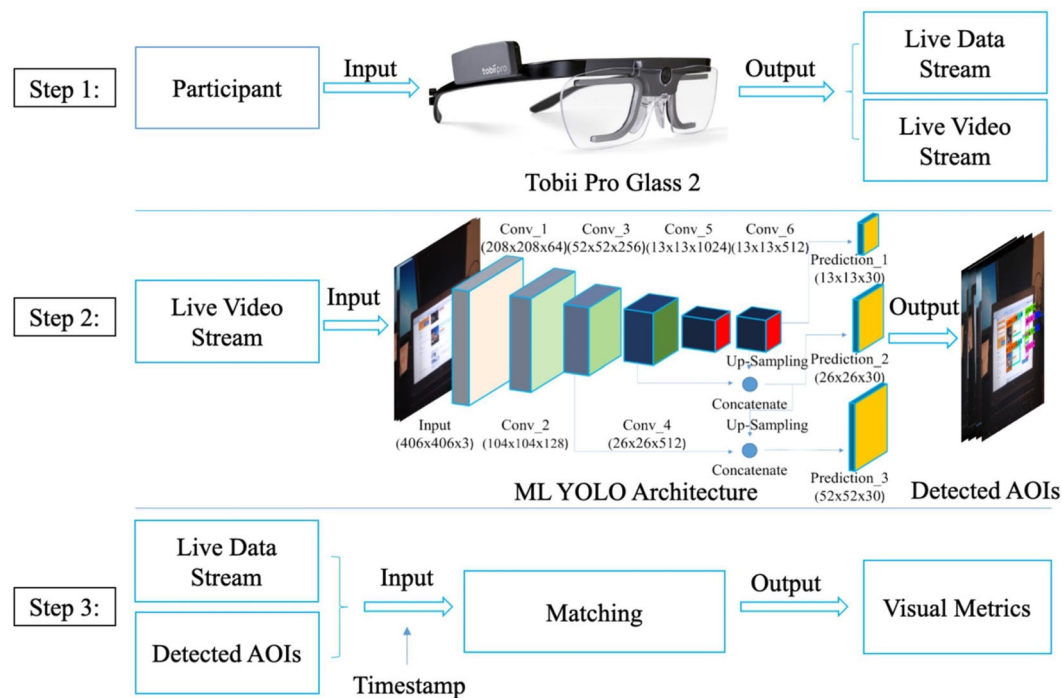


Figure 3. A framework of the proposed methodology.

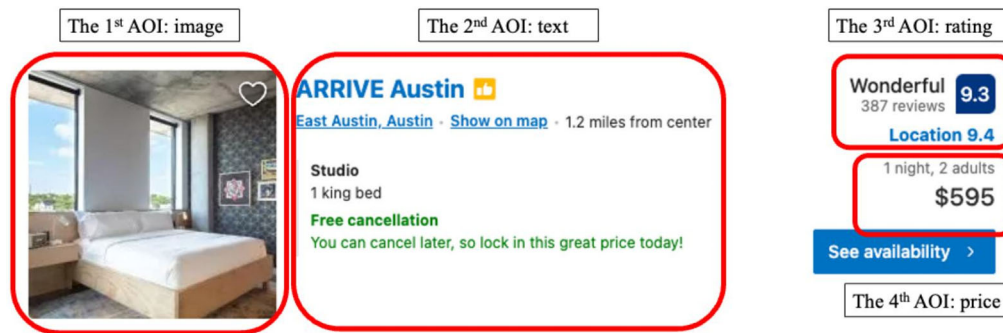


Figure 4. An example of four AOIs from a PC-based hotel advertisement.

setting took place in late 2021 (which had a two-year gap from the first one), some hotel brands may have changed their advertising information (e.g., hotel images). Such changes could have biased our results, but the bias was minimal because most participants viewed different hotels with distinct information across the two settings. To study consumers' visual attention to mobile ads, we focused on the same four AOIs, i.e., hotel image, rating, price, and text. We divided the text into two parts: the hotel name (second AOI in Figure 5) and other textual descriptions (fifth AOI in Figure 5) due to the unique layout of the Booking.com mobile app.

Detection of AOI Locations

With the rich recordings of consumers' visual focus during their online shopping sessions, we proposed

one- and two-step YOLO approaches to annotate the PC- and smartphone-based eye-tracking videos, respectively, in the following subsections.

One-Step YOLO

For the PC-based setting, we utilized the one-step YOLO to detect AOI locations from the 79 collected eye-tracking videos. The process involved three steps. First, we manually labeled the video frames for training and testing the YOLO model. As the Tobii Pro Glass 2 has a high sampling rate (up to 50 Hz), the information of consecutive frames is extremely similar in such a short time. Similar frames do not provide new information for the training. Hence, we randomly selected one frame out of every 10 from each video across all 79 videos. In our study, a two-minute-long video could generate about 3,000 frames. The 79 videos allowed us to generate approximately 40,000 initial

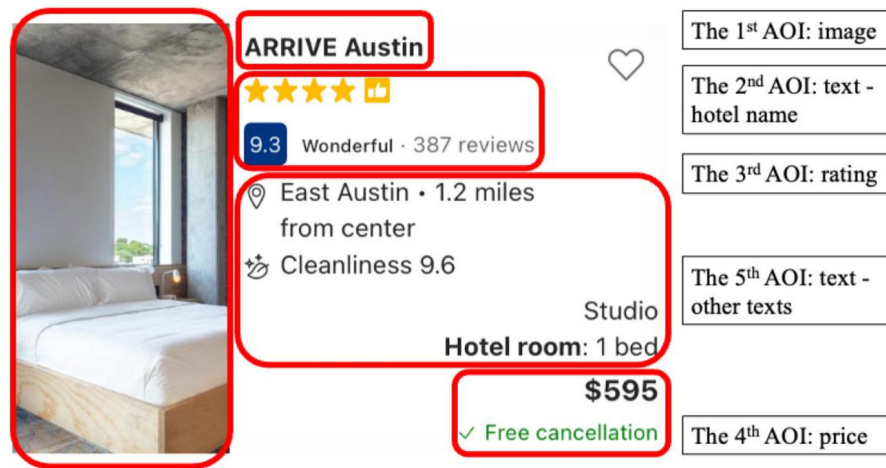


Figure 5. An example of five AOIs from a smartphone-based hotel advertisement.

selections. We manually identified and excluded about 1,200 low-quality frames because they were extremely vague. Low-quality frames can be ignored because they occurred mainly when participants were scrolling web pages up or down or moving their heads rapidly. During rapid movement, participants cannot gaze at any AOIs deliberately because human eyes cannot move as quickly as 30 motions per second (Al-Rahayfeh and Faezipour 2013), resulting in no record of eye fixation. As a result, we had approximately 38,800 high-quality frames from which we randomly sampled 8,000 frames for our final data set. During the labeling process, we manually annotated the AOIs' locations (i.e., hotel image, price, rating, and text) in each frame using YOLO Mark and obtained the annotated AOIs' coordinates and names. Afterward, we randomly divided the final dataset of 8,000 frames into three groups: one for training (6,000 frames), one for validation (800 frames), and one for testing (1,200 frames).

Second, we trained the one-step YOLO to detect AOIs from the labeled video frames. We evaluated the performance of the trained model using the validation data set after each training epoch. The training stopped once the validation accuracy did not increase in a certain number of consecutive epochs to avoid overfitting. Third, we assessed the performance of the trained one-step YOLO model using the testing data set. Given an input video frame, the outputs of the YOLO model were the detected AOIs' names and coordinates. With these outputs, we used the generic measure MAP (mean average precision) to evaluate the discrepancy between the AOI locations detected by the YOLO model and those identified by human coders. To this end, we first calculated a precision score and a recall score for one type of AOI (e.g.,

hotel image) using Equations (1) and (2) (Liu et al. 2020) below:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}, \quad (1)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}. \quad (2)$$

We classified the model-detected AOI as a true positive if its intersection over the union (IOU) with the corresponding manually labeled AOI was larger than a given threshold (0~1). Otherwise, we classified it as a false positive. We defined the manually labeled AOI as a false negative if the YOLO model failed to detect it. By varying the IOU threshold from 0 to 1 in classifying true and false positives, we could obtain a series of precision and recall scores.

Next, we computed the average precision (AP) in detecting one type of AOI, which is the area under the precision-recall curve, where the x-axis represents the recall scores, and the y-axis represents the precision scores (see Figure 6 for an illustration). Finally, we calculated the mean of the AP values across the four AOI types as follows:

$$\text{MAP} = \frac{AP_{\text{text}} + AP_{\text{price}} + AP_{\text{rating}} + AP_{\text{image}}}{4}. \quad (3)$$

Once the trained one-step YOLO achieves good performance with a high MAP, we can apply the model to detect the AOI locations in the PC-based eye-tracking videos in preparation to match them with consumers' eye fixations.

Two-Step YOLO

Detecting AOIs from smartphone-based eye-tracking videos is extremely challenging due to the small screens of smartphones, and thus, small-sized AOIs (Tong, Wu, and Zhou 2020). To overcome this

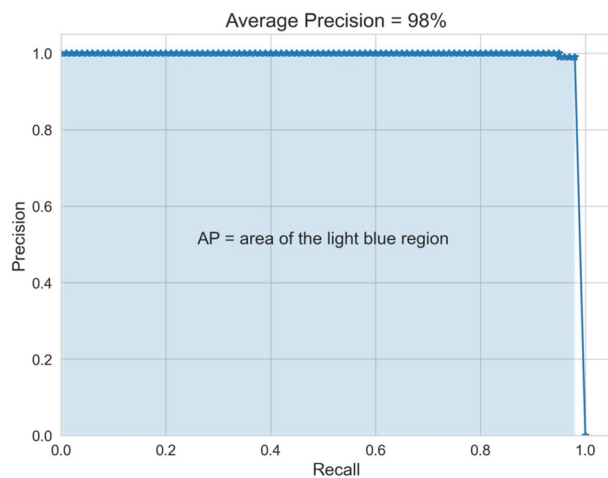


Figure 6. An example of the precision-recall curve and the average precision (AP).

challenge, we developed a two-step YOLO by extending the one-step YOLO, as we illustrate in Figure 7. The two-step YOLO involved two steps: (1) we first identified the smartphone's screen area in the video frames, which was an additional step to the one-step YOLO; (2) we then detected the AOIs on the screen area in the same way as the one-step YOLO. The two-step YOLO is a more advanced technique because it screens out noisy backgrounds outside the smartphone area to enlarge the AOIs' sizes, thereby improving the model performance and annotation accuracy.

We implemented the two-step YOLO approach to detect the AOIs' locations in the 148 smartphone-based eye-tracking videos through four steps. First, we obtained around 47,000 video frames by selecting one out of every 10 frames in the 148 videos. After excluding low-quality frames, we randomly sampled 1,400 high-quality frames to make our final data set. We then deployed the pre-trained YOLO v3 model with the MS COCO dataset (Redmon and Farhadi 2018) to detect the smartphone locations from the 1,400 frames. We randomly divided the final data set with the detected locations of the smartphones into training (1,080 frames), validation (144 frames), and testing (216 frames) data sets. The subsequent three processes—labeling, training, and evaluation—were the same as those in the one-step YOLO approach.

Match between Eye Fixations and Model-Detected AOIs

To obtain visual metrics, such as the eye-fixation count and duration, it was necessary to match the model-detected AOIs with consumers' eye-movement data. Accordingly, we conducted the matching as follows. The Tobii Pro Glass 2 helped us track raw eye-

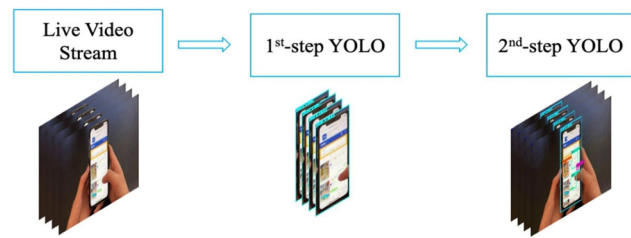


Figure 7. Two-step YOLO approach.

movement data (i.e., eye-gaze coordinates) and stored the data in a live data stream format, as we illustrate in Figure 3. The Tobii Pro Lab, a commercialized software provided by the Tobii company, then assisted us in generating consumers' eye-fixation coordinates with timestamps, indicating when and how long those eye-fixations occurred during the participants' online shopping session. We defined an eye fixation as a minimum fixation duration of 60 milliseconds, a default parameter in the Tobii Pro Lab. Then, we matched the eye-fixation coordinates with the YOLO model-detected AOIs frame by frame once the fixations were within the model-detected AOI. Upon completion of the matching, we calculated the participants' eye-fixation count and duration on each AOI during their entire shopping trajectory.

For illustration purposes, Figure 8 displays two instances of matching one participant's eye-fixation allocation with the detected AOIs in one video frame. As discussed previously, the eye-tracking technology allowed us to capture the participants' eye-fixation allocations, and our YOLO models identified each AOI's location from the video frame. Finally, we matched the eye fixation to one of the model-detected AOIs if the eye-fixation coordinates were within the AOI.

Data Analysis and Results

With the proposed methodology, we could collect participants' eye-fixation data and match them to each ad element in the PC and mobile ads during their online shopping sessions. In this section, we provide details of the variables and models used to test our hypotheses.

Variables

We collected 227 valid shopping videos (i.e., 79 on a PC and 148 on a smartphone). Each shopping video recorded a shopping session during which a participant completed booking a hotel in either Austin, TX, or Hawaii. Thus, our unit of analysis was a shopping

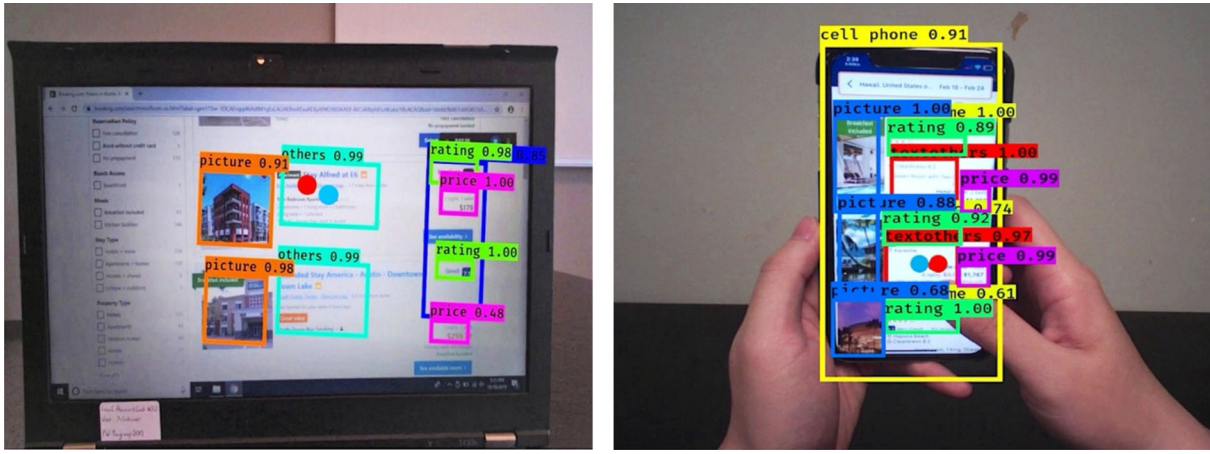


Figure 8. Screenshots of the matching processes. The left (right) panel shows a screenshot of matching the detected AOIs to participants' eye fixations in the PC-based (smartphone-based) experimental setting. The blue dot represents the center of the matched AOI, and the red dot indicates the participant's eye-fixation. The dark blue rectangle, representing the rating and price AOIs in the left panel, is used in training but does not affect the matching.

session, as we examined how frequently and how long a participant paid attention to each of the four AOIs (i.e., text, price, rating, and image) in each shopping session. Next, we define the variables used for further analyses.

Dependent Variables

To study how consumers allocate their attention to the ad elements in their shopping session, we operationalized the visual attention that each AOI received by the participants' eye-fixation count and duration on the respective AOI (Just and Carpenter 1980; Pieters and Wedel 2007). For each shopping session, we computed the eye-fixation count (*FC*) and duration (*FD*) allocated to each of the four AOIs (i.e., text, price, rating, and image).

Independent Variables

To test H1, the four AOIs served as our main independent variables: text, price, and rating for the textual ad elements, and image for the pictorial ones. Because the AOI is a categorical variable, we set image as the reference group and created three indicator variables for text, price, and rating, respectively: $AOI_k = 1$ for AOI k , and 0 otherwise, $k = \text{text, price, rating}$. To test hypothesis 2, we used an indicator variable for the shopping device, i.e., $Device_{mobile} = 1$ if a mobile device (i.e., smartphone) was used, and 0 if a PC was used. Last, to test hypothesis 3, we considered the interaction terms between each of the three indicator variables for the AOI (i.e., AOI_{rating} , AOI_{price} , AOI_{text}) and $Device_{mobile}$.

Control Variables

We included the following variables to control for possible confounding effects of the participants' age and gender: (1) $AOI_{18-25} = 1$ for participants between 18 and 25 years, and 0 for participants above 25 years. We used this binary variable because 88% of the participants fell into the two age groups, 18–25 years (53%) and 26–30 years (35%), with 9% in the group of 31–35 years, and 3% in the group of 36 years or older; (2) $Gender_{male} = 1$ for male participants and 0 for female participants. We added one variable to control for the potential effects of different travel destinations; (3) $Destination_{Hawaii} = 1$ for Hawaii and 0 for Austin, TX, and another to control for the effects of shopping time; (4) *Shopping Time* spent until finishing the shopping session (in minutes): the longer participants took to shop for hotels, the more they tended to pay attention to the ad elements. Table 2 provides descriptive statistics of the variables.

Regression Model

To examine how the ad element types (i.e., textual vs. pictorial), shopping device types (i.e., mobile vs. PC), and their interactions affect consumers' visual attention, we used a linear regression model specified by

$$\begin{aligned}
 Y_{ij} = & \alpha + \sum_{k=\text{text, price, rating}} \beta_{1,k} \times AOI_{kij} \\
 & + \beta_2 \times Device_{mobile, ij} \\
 & + \sum_{k=\text{text, price, rating}} \gamma_k \times AOI_{kij} \times Device_{mobile, ij} \\
 & + \delta \times X_{ij} + \varepsilon_{ij}.
 \end{aligned}
 \tag{4}$$

Table 2. Descriptive statistics.

Variable Type	Name	Min	Max	Mean	Std. Dev.
Dependent Variables	<i>FC</i> (eye-fixation Count)	0	8,335	724.82	995.40
	<i>FD</i> (eye-fixation Duration) (milliseconds)	0	324.12	29.60	39.58
Independent Variables	<i>Device_{mobile}</i>	0	1	0.65	0.48
	<i>AOI_{rating}</i>	0	1	0.25	0.43
	<i>AOI_{price}</i>	0	1	0.25	0.43
	<i>AOI_{text}</i>	0	1	0.25	0.43
Control Variables	<i>Age_{18–25}</i>	0	1	0.53	0.50
	<i>Gender_{male}</i>	0	1	0.74	0.44
	<i>Destination_{Hawaii}</i>	0	1	0.51	0.50
	<i>Shopping Time</i> (minutes)	0.24	11.66	2.56	1.98

Note: The number of observations is 908 (227 shopping sessions \times 4 AOIs)

The dependent variable Y_{ij} represents participant i 's eye-fixation count (FC) or duration (FD) on an AOI in shopping session j . AOI_k is the indicator variable for AOI k ($k = \text{text, price, or rating}$), and $Device_{mobile}$ is the indicator variable for mobile devices. X is a set of control variables, including Age_{18-25} , $Gender_{male}$, $Destination_{Hawaii}$, and $Shopping Time$. α is the intercept term; $\beta_{1,k}$ and β_2 are coefficients of AOI_k ($k = \text{text, price, or rating}$) and $Deice_{mobile}$, respectively; γ_k is the coefficient of the interaction term $AOI_k \times Device_{mobile}$ ($k = \text{text, price, or rating}$); δ are coefficients for the control variables; and ε is the random error term. We took the natural logarithmic transformation for the dependent variables (i.e., FC and FD) and one of the control variables, $Shopping Time$, to deal with their skewness. To test hypotheses 1 and 2, we used a model excluding the interaction terms (i.e., $AOI_k \times Device_{mobile}$) from Equation (4), while we utilized the model in Equation (4) to test hypothesis 3.

Significantly positive estimates of $\beta_{1,k}$ on FC (FD) imply that textual ad elements (i.e., text, price, and rating) receive more attention than pictorial ones (i.e., image), thus supporting hypothesis 1. Similarly, a significantly negative estimate of β_2 suggests that mobile ads receive less attention than PC ads, in line with hypothesis 2. Furthermore, positive, and significant, estimates of γ_k indicate that shopping on mobile devices strengthens the effect of textual ad elements receiving more attention than pictorial ones, which supports hypothesis 3.

In the following subsections, we first report the performance of the proposed YOLO models, and then present our findings on how consumers' visual attention varies across online ad elements (i.e., text, image, price, and rating) and shopping device types (i.e., mobile and PC).

Performance of the YOLO Model

Figure 9 provides the MAP values of the trained YOLO models. As illustrated in Panel A, the trained

YOLO model achieved a 94.5% MAP score across all four AOIs, compared to the precision score (i.e., presumably 100% correct after multiple trials) of the human coding in the PC-based experimental setting. In the smartphone-based setting, the trained YOLO model achieved an 86.8% MAP score over all AOIs, as shown in Panel B. These high MAP scores suggest that both of our proposed one- and two-step YOLO models showed high accuracy in detecting AOIs from the eye-tracking videos.

Additionally, our proposed YOLO models demonstrated high time efficiency. The human eye took about 30 s to manually identify all AOIs and match them with the participant's eye-fixation coordinates in one video frame. However, the trained two-step YOLO model took 0.2 s (150 times faster than human coding), and the trained one-step YOLO model took only 0.08 s (375 times faster than manual coding) to do the same. These results confirm that our proposed YOLO models were accurate and efficient in detecting AOIs in the eye-tracking videos collected from PCs and smartphones, ensuring the validity of our findings in the next subsections.

Consumers' Visual Attention across Different Ad Elements and Shopping Devices

Model-Free Evidence

By applying our proposed YOLO models, we obtained the eye-fixation count and duration on each AOI in each shopping session. We excluded potential bias in the results arising from the different-sized ad elements on screens between PCs and mobile devices, as we found similar proportions of each AOI on screens across the two devices (on average, 36.54% for hotel images, 8.24% for ratings, 46.65% for texts, and 8.57% for prices on the PC vs. 36.47% for hotel images, 9.03% for ratings, 45.56% for texts, and 8.93% for prices on the smartphone).

Figure 10 shows the boxplots of participants' eye-fixation counts allocated to the four AOIs on the PC

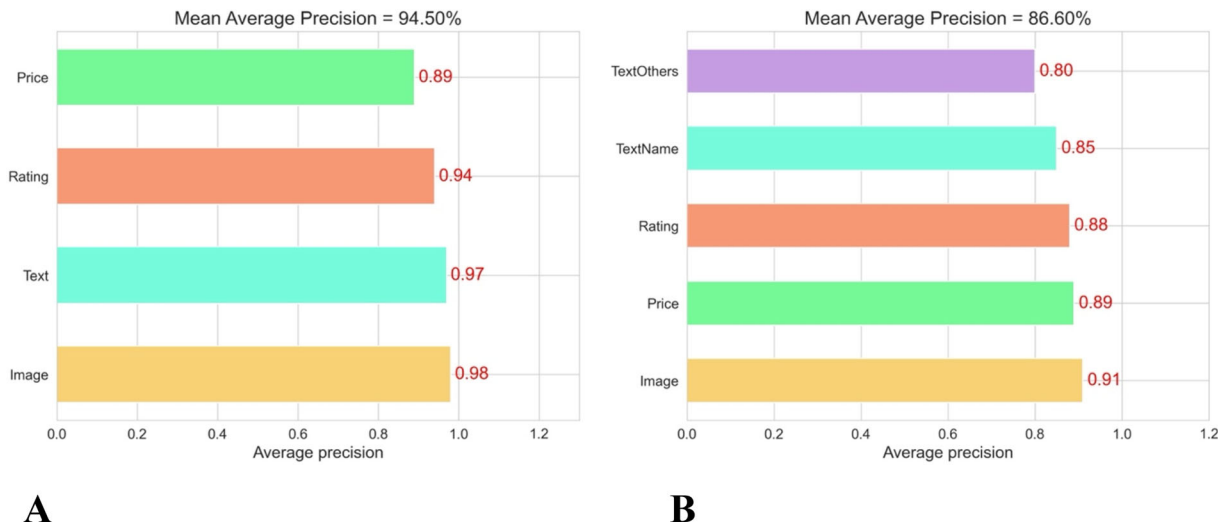


Figure 9. MAPs of the trained YOLO models in the two experimental settings: (A) PC-based setting; (B) smartphone-based setting.

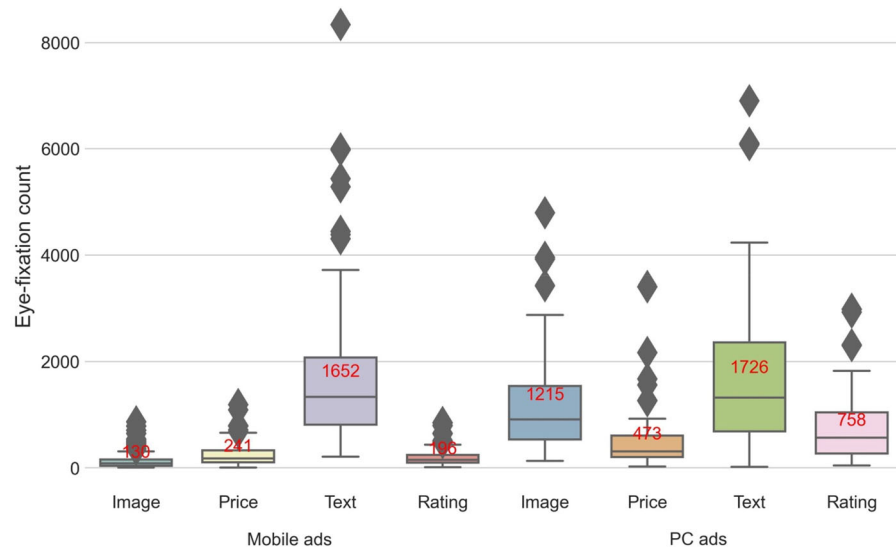


Figure 10. Boxplots of the eye-fixation counts on each AOI in PC and mobile ads. The numbers in red represent the mean eye-fixation counts over all shopping sessions. We note that the eye-fixation duration shows similar patterns (the results for the eye-fixation duration are available upon request).

and mobile ads. The results indicate that hotel texts captured more eye-fixation counts than hotel images, consistent across device types. While hotel images received more eye-fixation counts than hotel prices and ratings on PC ads, they received fewer eye-fixation counts than hotel prices and ratings on mobile ads. Moreover, compared to shopping on PCs, consumers allocated fewer eye-fixation counts to all four AOIs when shopping on mobile devices. We also found that these patterns were consistent with the eye-fixation duration. Thus, the preliminary results suggest that (1) textual ad elements receive more attention than pictorial ones, and such differences are more pronounced in ads on mobile devices than those on PCs; and (2) mobile ads receive less attention than

PC ads, which supports all hypotheses. We next report the regression analysis results to formally address our hypotheses.

Regression Analysis Results

Table 3 summarizes the estimation results of the regression model in Equation (4). We reported the Huber-White heteroscedasticity consistent standard errors to relax the assumption of homoscedastic residuals (White 1980). In the column “Main Effects Only,” the estimated coefficients for AOI_{rating} and AOI_{text} are positive and statistically significant; the coefficient for AOI_{price} is positive but statistically insignificant for both the eye-fixation count and duration. These results indicate that textual ad elements receive

Table 3. Estimation results of the effects of ad elements and device types on attention.

	Eye-fixation Count (FC)		Eye-fixation Duration (FD)	
	Main Effects Only	With Two-way Interactions	Main Effects Only	With Two-way Interactions
Intercept	4.739*** (0.106)	5.650*** (0.070)	1.507*** (0.101)	2.412*** (0.066)
AOI_{rating}	0.261** (0.076)	−0.560*** (0.084)	0.250** (0.074)	−0.544*** (0.080)
AOI_{price}	0.091 (0.088)	−1.039*** (0.076)	0.161 (0.085)	−1.014*** (0.074)
AOI_{text}	1.970*** (0.080)	0.277*** (0.067)	1.906*** (0.077)	0.256*** (0.064)
$Device_{mobile}$	−0.654*** (0.060)	−2.052*** (0.087)	−0.518*** (0.059)	−1.906*** (0.085)
$AOI_{rating} \times Device_{mobile}$		1.259*** (0.118)		1.217*** (0.113)
$AOI_{price} \times Device_{mobile}$		1.734*** (0.131)		1.803*** (0.126)
$AOI_{text} \times Device_{mobile}$		2.597*** (0.101)		2.530*** (0.096)
Age_{18-25}	0.101 (0.054)	0.101* (0.045)	0.087 (0.052)	0.087* (0.044)
$Gender_{male}$	0.150* (0.066)	0.150** (0.054)	0.166** (0.062)	0.166** (0.049)
$Destination_{Hawaii}$	−0.043 (0.053)	−0.043 (0.043)	−0.040 (0.050)	−0.040 (0.041)
$Shopping\ Time\ (minutes)$	1.052*** (0.040)	1.052*** (0.031)	1.046*** (0.039)	1.046*** (0.029)

Note: *** $p < 0.001$, ** $p < .01$, * $p < .05$. Huber-White heteroscedasticity consistent standard errors are in parentheses.

more attention than pictorial ones. Thus, this finding supports hypothesis 1. Additionally, the significantly negative coefficients for $Device_{mobile}$ confirm that mobile ads receive less attention than PC ads, supporting hypothesis 2. In the column “With Two-Way Interactions,” the three estimated coefficients for the interaction terms, $AOI_{rating, price, or text} \times Device_{mobile}$, are positive and statistically significant. This finding implies that, when consumers shop on mobile devices, it strengthens the effect that textual ad elements receive more attention than pictorial ones in ads on mobile devices, which supports hypothesis 3.

Regarding the control variables, consumers who were younger, male, and who spent more time shopping tended to pay more attention to hotel ads, while the destination seemed not to have much influence on their visual attention. These findings are consistent across the two dependent variables and model specifications.

Additional Analyses

Consumers exhibit varying search and decision-making strategies while shopping online (Pan, Zhang, and Law 2013). Thus, it is important to study whether and how their attention changes over time during their shopping journey. Existing studies have reported how consumers search for information in different settings. For example, Haan and Moraga-González (2011) found that consumers tend to visit brands with the most salient advertising when searching for products. Ghose, Goldfarb, and Han (2013) discussed how consumer search behavior on mobile devices can be impacted by the limited display capability on smaller screens, resulting in increased costs of information gathering, compared to PCs. Zhu and Dukes (2017) demonstrated how firms can selectively promote product attributes to manipulate consumers’ limited attention during the search and evaluation process. Our

study contributes to this research stream on consumer search behavior by (1) introducing the YOLO methods such that researchers can collect and annotate granular eye-tracking data to examine the dynamics of consumers’ visual attention over time, and (2) adding empirical evidence of how consumers’ visual attention changes from their search for products to their decisions on purchasing products while shopping online.

To this end, we collected additional information on the shopping progress level (in proportion to the percentage of the entire shopping time) by median, splitting each shopping session into two subgroups: one from 0 to 50% and the other from 50% to 100%. As we matched eye fixations to the model-detected AOIs using our proposed YOLO approaches, video frame by video frame, we computed each participant’s eye-fixation count and duration on each AOI for each shopping progress group. Thus, we obtained 1,816 observations (227 shopping sessions \times 4 AOIs \times 2 shopping progress groups). To measure the effects of the shopping progress level on visual attention, we used the following linear regression model:

$$Y_{ilj} = \alpha + \theta \times Shopping\ Progress_{50\%-100\%,\ ij} + \sum_{k=text, price, rating} \beta_{1,k} \times AOI_{kilj} + \beta_2 \times Device_{mobile,\ ij} + \delta \times X_{ij} + \varepsilon_{ilj}, \quad (5)$$

where i indexes the participant, and l represents the shopping progress level (i.e., 0–50% and 50–100%) in shopping session j . The variable $Shopping\ Progress_{50\%-100\%,\ ij}$ is an indicator variable for the shopping progress level (= 1 for 50–100%; = 0 for 0–50%), and the corresponding coefficient θ is our main interest. If the estimated θ is significantly positive, then consumers pay more attention to ads in the shopping progress level of 50–100%, compared to that of 0–50%. The other variables and parameters in Equation (5) have the same definitions and interpretations as those in Equation (4).

Table 4. Visual attention changes during the online shopping session.

	Eye-fixation Count (FC)	Eye-fixation Duration (FD)
Intercept	3.930*** (0.084)	0.719*** (0.079)
<i>Shopping Progress</i> _{50%–100%}	0.100* (0.039)	0.115** (0.037)
<i>AOI</i> _{rating}	0.266*** (0.058)	0.240*** (0.055)
<i>AOI</i> _{price}	0.078 (0.078)	0.150* (0.062)
<i>AOI</i> _{text}	1.997*** (0.059)	1.916*** (0.056)
<i>Device</i> _{mobile}	−0.629*** (0.048)	−0.492*** (0.046)
<i>Age</i> _{18–25}	0.109** (0.041)	0.092* (0.039)
<i>Gender</i> _{male}	0.138** (0.048)	0.156** (0.045)
<i>Destination</i> _{Hawaii}	−0.037 (0.040)	−0.029 (0.037)
<i>Shopping Time</i> (minutes)	1.078*** (0.032)	1.053*** (0.030)

Note: *** $p < 0.001$, ** $p < .01$, * $p < .05$. Huber-White heteroscedasticity consistent standard errors are in parentheses.

We reported the estimation results of Equation (5) in Table 4. After controlling for the effects of ad elements, device types, and other covariates, we found a significantly positive coefficient for the variable *Shopping Progress*_{50%–100%} for both dependent variables (eye-fixation count and duration). These results indicate that hotel ads receive more attention when consumers are finishing their hotel searches and deciding on hotel choices, compared to when they search for hotels earlier in their shopping trajectory. Such attentional changes suggest that consumers adopt different search and decision-making strategies while shopping online for hotels. With many hotel options available, consumers are likely to first glance at the options and form a consideration set for further scrutiny (Noone and Robson 2016). Afterward, they gather more information and deliberate the details of each option for the final selection. Because the deliberation process involves more information acquisition (Noone and Robson 2016), consumers pay more attention to ads when they are close to making their final decision, corroborating our empirical findings.

Discussion

With the prevalence of mobile devices as an advertising channel, it is important to understand consumers' visual attention in mobile advertisements, as it can potentially offer new insights regarding online advertising strategies. However, few eye-tracking studies on mobile advertising exist mainly due to technical difficulties such as the time-consuming manual annotation of eye-tracking data (Meißner et al. 2019) and the detection of small AOIs on small mobile screens (Tong, Wu, and Zhou 2020). Thus, automatic annotations that can precisely detect small AOIs are necessary, but few reliable algorithms exist in advertising or marketing research (Meißner et al. 2019). To overcome these challenges, we applied an ML-based

object-detection technique, YOLO v3, to automatically annotate the eye-tracking data. Specifically, we proposed a one- and two-step YOLO to process the eye-tracking videos collected from PCs and smartphones, respectively. With the proposed YOLO approaches, we further probed into how ad element types (i.e., textual and pictorial), shopping device types (i.e., mobile and PC), and their interactions affected consumers' visual attention while shopping online.

Theoretical and Methodological Contributions

This study contributes to advertising research in three ways. First, we contribute to the literature on the theory of attention. Consumers' visual attention to online advertisements is affected by two types of factors: bottom-up and top-down (James 1890). The salience of visual stimuli affects bottom-up attention, while the informativeness of these stimuli affects top-down attention (Chun and Wolfe 2005; Pieters and Wedel 2007). In this study, we empirically explored consumers' visual attention across ad element types and shopping devices, which influence the salience and informativeness of ad elements.

Consistent with hypothesis 1, we found that textual ad elements received more attention than pictorial ad elements, implying that textual ad elements are more salient and informative than pictorial ones. Although pictorial ad elements are typically more salient (Pieters and Wedel 2007), we speculate that the smaller sizes of pictorial ad elements (images), as compared to textual ones in our empirical setting, i.e., online hotel advertisements (see Figures 4 and 5), can make pictorial ad elements less salient, and thus less likely to capture attention (Janiszewski 1998). Furthermore, our finding corroborates the fact that textual ad elements are more informative than pictorial ones when consumers shop online (Rayner et al. 2001). These findings contribute to the theory of attention by identifying the effects of various stimuli (i.e., textual and pictorial ad elements) on visual attention.

Moreover, we found that mobile ads received less attention than PC ads, which is in line with hypothesis 2. This finding extends our understanding of consumers' attentional differences in ads across two major online shopping devices, i.e., PCs and mobile devices. Mobile ads are smaller than PC ads due to the smaller and narrower screens of mobile devices (Ghose, Goldfarb, and Han 2013; Wang et al. 2019), making mobile ads less salient (Janiszewski 1998) and leading to less attentional demand on consumers to process mobile ads (Pieters and Wedel 2004). Because

the salience and informativeness of the same ad elements may vary across PC and mobile ads, we further studied the interaction effects between ad element and device types. Our results show that shopping on mobile devices strengthens the effect that textual ad elements receive more attention than pictorial ones, thus supporting hypothesis 3. The rectangular layout of pictorial ad elements on mobile devices makes them not only less salient but also less informative than those with a square layout on PCs (see [Figures 4 and 5](#)) ([Krider, Raghubir, and Krishna 2001](#)): the rectangular layout may hide much of the pictorial ad elements' content due to the shorter horizontal space on mobile screens. These findings enhance our knowledge of cross-device effects regarding consumers' visual attention.

Second, our proposed two-step YOLO, which can efficiently and accurately annotate ad elements in eye-tracking data collected from smartphones, makes it easier for researchers to analyze small and complex visual stimuli on mobile devices. This approach contributes to a growing demand for eye-tracking studies with mobile devices (e.g., [Pozharliev, Rossi, and Angelis 2022](#)), as we demonstrated the high accuracy and efficiency of the proposed YOLO approaches in detecting and annotating AOIs in videos collected with a portable eye-tracking device. In the PC-based experimental setting, it took about 66 h for us to manually label the 8,000 frames for training and testing the one-step YOLO, resulting in as much as 3,542 h for all 425,000 frames. However, the trained one-step YOLO took approximately nine hours to finish the same task. Thus, our proposed YOLO methods can significantly reduce labor costs and offer a reliable automated object-detection algorithm for eye-tracking research. As such, we expect to see a broader and deeper research scope in eye-tracking studies in advertising, media, and marketing research.

Third, our study contributes to the consumer search literature. Consumers tend to adopt different search strategies on their path to purchasing products ([Zhu and Dukes 2017](#)), thus affecting their visual attention to online ads. Online shoppers typically first search for products to generate a consideration set, and then deliberately evaluate each product before making a final choice ([Noone and Robson 2016](#)). Then, consumers pay more attention to ads when they proceed to the deliberation process of decision making. Analogous to this process from consumer searches to decision making on purchases, we revealed that online ads receive more visual attention when consumers are close to deciding on hotel choices, compared to when they search for hotels earlier in their shopping trajectory.

Practical Implications

This research offers the following practical insights. First, our findings can help advertising platforms deliver their marketing messages more effectively. Our results suggest that textual ad elements receive more attention than pictorial ones; thus, online advertising platforms could strategically display important messages using textual formats instead of pictorial ones. Second, our findings suggest that customized layouts of online ads for different shopping devices, e.g., PCs and mobile apps, can improve effectiveness. Currently, major online travel agencies (e.g., [Expedia.com](#) and [Booking.com](#)) uniformly use rectangular layouts to display hotel images on mobile apps, but square-shaped layouts on PCs. However, according to our results, the difference between PC and mobile ad layouts can decrease the effectiveness of mobile ads: the rectangular layout on mobile devices can lead to a loss of salience and informativeness with respect to pictorial ad elements. Thus, we suggest that online advertisers be aware of such phenomena to develop more effective cross-device strategies to improve consumers' attention to their ads. Finally, practitioners should consider both the salience and informativeness of an ad element when designing ad layouts, as they both play significant roles in capturing consumers' visual attention.

Limitations and Suggestions for Future Research

This research has the following limitations. First, the participants in our experiment were all university students. Though we recruited participants randomly, more male than female subjects participated due to the imbalanced population in the school where we conducted our experiment. Thus, our results may not be fully generalizable to broader consumer groups. Second, due to the pandemic, there was a two-year gap between the two experimental settings. As such, some hotels may have changed their advertising information (e.g., images and prices) between the two periods. This gap could have biased our results, as such changes can influence consumers' online shopping behavior. Third, we conducted the experiment in a university lab, which may differ from the participants' actual shopping environments and may have impacted their shopping behaviors. Last, when we conducted the PC-based experimental setting, YOLO v3 was the state-of-the-art object-detection model, while an advanced version, YOLO v7, is available now. Nevertheless, we argue that technological advances cannot devalue our methodological contribution to advertising research because researchers can easily

adapt our proposed YOLO approaches to other object-detection methods, including YOLO v7.

Our proposed YOLO approaches, combined with eye-tracking technology, offer many opportunities for future researchers and practitioners. We thus conclude by listing several fruitful research areas. One possible research direction would involve applying our framework to obtain more complicated dynamic visual metrics. For example, with effortless modifications in our programming code, researchers could record each consumer's viewing trajectory across different AOIs. This process could address more complex problems concerning temporal changes in consumers' visual attention in advertising research. Second, researchers could deploy our proposed YOLO approaches to study general video advertisements (e.g., livestream videos). The YOLO approaches can detect various ad elements (e.g., brand logos, product images, and influencers) and annotate their locations in videos, allowing researchers to probe into how the locations of ad elements impact video ads' effectiveness. Third, researchers can use the YOLO model to analyze real-world eye-tracking data obtained from actual retail settings and other online shopping platforms, which benefits from annotating the complex locations of dynamic AOIs efficiently. Last, the YOLO approaches can be applied to extract facial and gender features from human subjects in eye-tracking data, providing opportunities for interdisciplinary research such as diversity, equity, and inclusion in online shopping environments.

Ethics Statement

This work received the approval of the Institutional Review Board (IRB) at the University of Houston (STUDY0 0001635).

Disclosure Statement

The authors have no conflicts of interest to disclose.

Funding

This work was supported by 2021 Amazon Research Award.

ORCID

Wen Xie  <http://orcid.org/0000-0002-0274-4981>
 Mi Hyun Lee  <http://orcid.org/0000-0001-9747-2888>
 Ming Chen  <http://orcid.org/0009-0004-2860-6749>
 Zhu Han  <http://orcid.org/0000-0002-6606-5822>

References

- Al-Rahayfeh, A., and M. Faezipour. 2013. "Enhanced Frame Rate for Real-Time Eye Tracking Using Circular Hough Transform." 2013 IEEE Long Island Systems, Applications and Technology Conference, New York.
- Barz, M., and D. Sonntag. 2021. "Automatic Visual Attention Detection for Mobile Eye Tracking Using Pre-Trained Computer Vision Models and Human Gaze." *Sensors* 21 (12): 4143. <https://doi.org/10.3390/s21124143>
- Beuckels, E., L. Hudders, V. Cauberghe, K. Bombeke, W. Durnez, and J. Morton. 2021. "To Fit in or to Stand out? An Eye-Tracking Study Investigating Online Banner Effectiveness in a Media Multitasking Context." *Journal of Advertising* 50 (4): 461–478. <https://doi.org/10.1080/00913367.2020.1870053>
- Brône, G., B. Oben, and T. Goedemé. 2011. "Towards a More Effective Method for Analyzing Mobile Eye-tracking Data: Integrating Gaze Data with Object Recognition Algorithms." Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-based Interaction, Beijing, China.
- Callemein, T., K. V. Beeck, G. Brône, and T. Goedemé. 2018. "Automated Analysis of Eye-Tracker-Based Human-Human Interaction Studies." International Conference on Information Science and Applications, Singapore.
- Cao, Z., G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh. 2021. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (1): 172–186. <https://doi.org/10.1109/TPAMI.2019.2929257>
- Carter, B. T., and S. G. Luke. 2020. "Best Practices in Eye Tracking Research." *International Journal of Psychophysiology* 155: 49–62. <https://doi.org/10.1016/j.ijpsycho.2020.05.010>
- Chen, M., R. R. Burke, S. K. Hui, and A. Leykin. 2021. "Understanding Lateral and Vertical Biases in Consumer Attention: An in-Store Ambulatory Eye-Tracking Study." *Journal of Marketing Research* 58 (6): 1120–1141. <https://doi.org/10.1177/002224372199837>
- Chevalier, S. 2022. "U.S. Online Retail Website Visits and Orders 2022, by Device." Statista, July 21. <https://www.statista.com/statistics/201680/retail-site-device-visit-order-share-usa/>
- Chun, M., and J. Wolfe. 2005. "Visual Attention." In *Blackwell Handbook of Sensation and Perception*, edited by E. Bruce Goldstein, 272–310. Malden, MA: Blackwell Publishing.
- Ghose, A., A. Goldfarb, and S. P. Han. 2013. "How Is the Mobile Internet Different? Search Costs and Local Activities." *Information Systems Research* 24 (3): 613–631. <https://doi.org/10.1287/isre.1120.0453>
- Haan, M. A., and J. L. Moraga-González. 2011. "Advertising for Attention in a Consumer Search Model." *The Economic Journal* 121 (552): 552–579. <https://doi.org/10.1111/j.1468-0297.2011.02423.x>
- Hernández-Méndez, J., and F. Muñoz-Leiva. 2015. "What Type of Online Advertising is Most Effective for eTourism 2.0? An Eye Tracking Study Based on the Characteristics of Tourists." *Computers in Human*

- Behavior* 50: 618–625. <https://doi.org/10.1016/j.chb.2015.03.017>
- Hessels, R. S., J. S. Benjamins, T. H. W. Cornelissen, and I. T. C. Hooge. 2018. “A Validation of Automatically-Generated Areas-of-Interest in Videos of a Face for Eye-Tracking Research.” *Frontiers in Psychology* 9: 1367. <https://doi.org/10.3389/fpsyg.2018.01367>
- IAB. 2023. “Internet Advertising Revenue Report: Full Year 2022.” *Interactive Advertising Bureau (IAB)*, April 12. <https://www.iab.com/insights/internet-advertising-revenue-report-full-year-2022/>
- James, W. 1890. *The Principles of Psychology*. New York: H. Holt and Company.
- Janiszewski, C. 1998. “The Influence of Display Characteristics on Visual Exploratory Search Behavior.” *Journal of Consumer Research* 25 (3): 290–301. <https://doi.org/10.1086/209540>
- Jayawardena, G., and S. Jayarathna. 2021. “Automated Filtering of Eye Movements Using Dynamic AOI in Multiple Granularity Levels.” *International Journal of Multimedia Data Engineering and Management* 12 (1): 49–64. <https://doi.org/10.4018/IJMDem.2021010104>
- Jongerius, C., T. Callemeyn, T. Goedemé, K. Van Beeck, J. A. Romijn, E. M. A. Smets, and M. A. Hillen. 2021. “Eye-Tracking Glasses in Face-to-Face Interactions: Manual versus Automated Assessment of Areas-of-Interest.” *Behavior Research Methods* 53 (5): 2037–2048. <https://doi.org/10.3758/s13428-021-01544-2>
- Just, M. A., and P. A. Carpenter. 1980. “A Theory of Reading: From Eye Fixations to Comprehension.” *Psychological Review* 87 (4): 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Krider, R. E., P. Raghubir, and A. Krishna. 2001. “Pizzas: π or Square? Psychophysical Biases in Area Comparisons.” *Marketing Science* 20 (4): 405–425. <https://doi.org/10.1287/mksc.20.4.405.9756>
- Kurzahls, K., M. Hlawatsch, C. Seeger, and D. Weiskopf. 2017. “Visual Analytics for Mobile Eye Tracking.” *IEEE Transactions on Visualization and Computer Graphics* 23 (1): 301–310. <https://doi.org/10.1109/TVCG.2016.2598695>
- Liu, L., W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen. 2020. “Deep Learning for Generic Object Detection: A Survey.” *International Journal of Computer Vision* 128 (2): 261–318. <https://doi.org/10.1007/s11263-019-01247-4>
- Lohse, G. L. 1997. “Consumer Eye Movement Patterns on Yellow Pages Advertising.” *Journal of Advertising* 26 (1): 61–73. <https://doi.org/10.1080/00913367.1997.10673518>
- Maslowska, E., C. M. Segijn, K. A. Vakeel, and V. Viswanathan. 2020. “How Consumers Attend to Online Reviews: An Eye-Tracking and Network Analysis Approach.” *International Journal of Advertising* 39 (2): 282–306. <https://doi.org/10.1080/02650487.2019.1617651>
- Meißner, M., J. Pfeiffer, T. Pfeiffer, and H. Oppewal. 2019. “Combining Virtual Reality and Mobile Eye Tracking to Provide a Naturalistic Experimental Environment for Shopper Research.” *Journal of Business Research* 100: 445–458. <https://doi.org/10.1016/j.jbusres.2017.09.028>
- Myers, S. D., G. D. Deitz, B. A. Huhmann, S. Jha, and J. H. Tata. 2020. “An Eye-Tracking Study of Attention to Brand-Identifying Content and Recall of Taboo Advertising.” *Journal of Business Research* 111: 176–186. <https://doi.org/10.1016/j.jbusres.2019.08.009>
- Noone, B. M., and S. K. Robson. 2016. “Understanding Consumers’ Inferences from Price and Nonprice Information in the Online Lodging Purchase Decision.” *Service Science* 8 (2): 108–123. <https://doi.org/10.1287/serv.2016.0141>
- Orquin, J. L., and S. M. Loose. 2013. “Attention and Choice: A Review on Eye Movements in Decision Making.” *Acta Psychologica* 144 (1): 190–206. <https://doi.org/10.1016/j.actpsy.2013.06.003>
- Pan, B., L. Zhang, and R. Law. 2013. “The Complex Matter of Online Hotel Choice.” *Cornell Hospitality Quarterly* 54 (1): 74–83. <https://doi.org/10.1177/19389655124632>
- Peschel, A. O., and J. L. Orquin. 2013. “A Review of the Findings and Theories on Surface Size Effects on Visual Attention.” *Frontiers in Psychology* 4: 902. <https://doi.org/10.3389/fpsyg.2013.00902>
- Pew Research Center. 2021. “Mobile Fact Sheet.” Pew Research Center, April 07. <https://www.pewresearch.org/internet/fact-sheet/mobile/>
- Pfeiffer, J., T. Pfeiffer, M. Meißner, and E. Weiß. 2020. “Eye-Tracking-Based Classification of Information Search Behavior Using Machine Learning: Evidence from Experiments in Physical Shops and Virtual Reality Shopping Environments.” *Information Systems Research* 31 (3): 675–691. <https://doi.org/10.1287/isre.2019.0907>
- Pfiffelmann, J., N. Dens, and S. Soulez. 2020. “Personalized Advertisements with Integration of Names and Photographs: An Eye-Tracking Experiment.” *Journal of Business Research* 111: 196–207. <https://doi.org/10.1016/j.jbusres.2019.08.017>
- Pieters, R., and M. Wedel. 2004. “Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects.” *Journal of Marketing* 68 (2): 36–50. <https://doi.org/10.1509/jmkg.68.2.36.27794>
- Pieters, R., and M. Wedel. 2007. “Goal Control of Attention to Advertising: The Yarus Implication.” *Journal of Consumer Research* 34 (2): 224–233. <https://doi.org/10.1086/519150>
- Pieters, R., M. Wedel, and R. Batra. 2010. “The Stopping Power of Advertising: Measures and Effects of Visual Complexity.” *Journal of Marketing* 74 (5): 48–60. <https://doi.org/10.1509/jmkg.74.5.0>
- Pozharliev, R., D. Rossi, and M. D. Angelis. 2022. “A Picture Says More Than a Thousand Words: Using Consumer Neuroscience to Study Instagram Users’ Responses to Influencer Advertising.” *Psychology & Marketing* 39 (7): 1336–1349. <https://doi.org/10.1002/mar.21659>
- Rayner, K., B. Miller, and C. M. Rotello. 2008. “Eye Movements When Looking at Print Advertisements: The Goal of the Viewer Matters.” *Applied Cognitive Psychology* 22 (5): 697–707. <https://doi.org/10.1002/acp.1389>
- Rayner, K., C. M. Rotello, A. J. Stewart, J. Keir, and S. A. Duffy. 2001. “Integrating Text and Pictorial Information: Eye Movements When Looking at Print Advertisements.” *Journal of Experimental Psychology. Applied* 7 (3): 219–226. <https://doi.org/10.1037/1076-898X.7.3.219>

- Redmon, J., and A. Farhadi. 2018. "YOLOv3: An Incremental Improvement." arXiv. <https://doi.org/10.48550/ARXIV.1804.02767>
- Schmidt, L. L., and E. Maier. 2022. "Interactive Ad Avoidance on Mobile Phones." *Journal of Advertising* 51 (4): 440–449. <https://doi.org/10.1080/00913367.2022.2077266>
- Scott, N., R. Zhang, D. Le, and B. Moyle. 2019. "A Review of Eye-Tracking Research in Tourism." *Current Issues in Tourism* 22 (10): 1244–1261. <https://doi.org/10.1080/13683500.2017.1367367>
- Segijn, C. M., H. A. Voorveld, and K. A. Vakeel. 2021. "The Role of Ad Sequence and Privacy Concerns in Personalized Advertising: An Eye-Tracking Study into Synced Advertising Effects." *Journal of Advertising* 50 (3): 320–329. <https://doi.org/10.1080/00913367.2020.1870586>
- Shi, S. W., M. Wedel, and F. G. M. Pieters. 2013. "Information Acquisition during Online Decision Making: A Model-Based Exploration Using Eye-Tracking Data." *Management Science* 59 (5): 1009–1026. <https://doi.org/10.1287/mnsc.1120.1625>
- Simola, J., J. Kuisma, and J. K. Kaakinen. 2020. "Attention, Memory, and Preference for Direct and Indirect Print Advertisements." *Journal of Business Research* 111: 249–261. <https://doi.org/10.1016/j.jbusres.2019.06.028>
- Tong, K., Y. Wu, and F. Zhou. 2020. "Recent Advances in Small Object Detection Based on Deep Learning: A Review." *Image and Vision Computing* 97: 103910. <https://doi.org/10.1016/j.imavis.2020.103910>
- Van der Lans, R., R. Pieters, and M. Wedel. 2008. "Research Note—Competitive Brand Salience." *Marketing Science* 27 (5): 922–931. <https://doi.org/10.1287/mksc.1070.0327>
- Wang, F., L. Zuo, Z. Yang, and Y. Wu. 2019. "Mobile Searching versus Online Searching: Differential Effects of Paid Search Keywords on Direct and Indirect Sales." *Journal of the Academy of Marketing Science* 47 (6): 1151–1165. <https://doi.org/10.1007/s11747-019-00649-7>
- Wedel, M., and R. Pieters. 2006. "Eye Tracking for Visual Marketing." *Foundations and Trends® in Marketing* 1 (4): 231–320. <https://doi.org/10.1561/17000000011>
- White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–838. <https://doi.org/10.2307/1912934>
- Wooley, B., S. Bellman, N. Hartnett, A. Rask, and D. Varan. 2022. "Influence of Dynamic Content on Visual Attention during Video Advertisements." *European Journal of Marketing* 56 (13): 137–166. <https://doi.org/10.1108/EJM-10-2020-0764>
- Yantis, S., and J. Jonides. 1990. "Abrupt Visual Onsets and Selective Attention: Voluntary versus Automatic Allocation." *Journal of Experimental Psychology. Human Perception and Performance* 16 (1): 121–134. <https://doi.org/10.1037//0096-1523.16.1.121>
- Ye, Q., and D. Doermann. 2015. "Text Detection and Recognition in Imagery: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (7): 1480–1500. <https://doi.org/10.1109/TPAMI.2014.2366765>
- Zhu, Y., and A. Dukes. 2017. "Prominent Attributes under Limited Attention." *Marketing Science* 36 (5): 683–698. <https://doi.org/10.1287/mksc.2017.1037>