

Influence of dynamic content on visual attention during video advertisements

Influence of
dynamic
content

137

Brooke Wooley

Murdoch University, Perth, Australia

Steven Bellman and Nicole Hartnett

*Ehrenberg-Bass Institute for Marketing Science, University of South Australia,
Adelaide, Australia, and*

Amy Rask and Duane Varan

MediaScience, Austin, Texas, USA

Received 19 October 2020

Revised 4 June 2021

2 December 2021

4 May 2022

Accepted 15 May 2022

Abstract

Purpose – Dynamic advertising, including television and online video ads, demands new theory and tools developed to understand attention to moving stimuli. The purpose of this study is to empirically test the predictions of a new dynamic attention theory, Dynamic Human-Centred Communication Systems Theory, versus the predictions of salience theory.

Design/methodology/approach – An eye-tracking study used a sample of consumers to measure visual attention to potential areas of interest (AOIs) in a random selection of unfamiliar video ads. An eye-tracking software feature called intelligent bounding boxes (IBBs) was used to track attention to moving AOIs. AOIs were coded for the presence of static salience variables (size, brightness, colour and clutter) and dynamic attention theory dimensions (imminence, motivational relevance, task relevance and stability).

Findings – Static salience variables contributed 90% of explained variance in fixation and 57% in fixation duration. However, the data further supported the three-way interaction uniquely predicted by dynamic attention theory: between imminence (central vs peripheral), relevance (motivational or task relevant vs not) and stability (fleeting vs stable). The findings of this study indicate that viewers treat dynamic stimuli like real life, paying less attention to central, relevant and stable AOIs, which are available across time and space in the environment and so do not need to be memorised.

Research limitations/implications – Despite the limitations of small samples of consumers and video ads, the results of this study demonstrate the potential of two relatively recent innovations, which have received limited emphasis in the marketing literature: dynamic attention theory and IBBs.

Practical implications – This study documents what does and does not attract attention to video advertising. What gets attention according to salience theory (e.g. central location) may not always get attention in dynamic advertising because of the effects of relevance and stability. To better understand how to execute video advertising to direct and retain attention to important AOIs, advertisers and advertising researchers are encouraged to use IBBs.

Originality/value – This study makes two original contributions: to marketing theory, by showing how dynamic attention theory can predict attention to video advertising better than salience theory, and to



© Brooke Wooley, Steven Bellman, Nicole Hartnett, Amy Rask and Duane Varan. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

European Journal of Marketing

Vol. 56 No. 13, 2022

pp. 137-166

Emerald Publishing Limited

0309-0566

DOI 10.1108/EJM-10-2020-0764

marketing research, showing the utility of tracking visual attention to moving objects in video advertising with IBBs, which appear underutilised in advertising research.

Keywords Video advertising, Visual attention, Eye-tracking, Saliency, Dynamic human-centric communication systems theory, Television advertising

Paper type Research paper

Introduction

Advertising expenditure has steadily shifted over time from static media (e.g. print, billboards and online display ads) to dynamic media (e.g. television and online video ads). Television and online video are forecast to jointly attract around 40% of global advertising media spend in 2022 (Dentsu, 2021). Attention is one construct advertisers use to assess advertising effectiveness, including for dynamic ads, as a precursor to other downstream effects such as recall and behaviour. There are numerous approaches to measure attention, but a most direct and popular method is to measure visual attention with eye-tracking.

Theories about what draws attention to advertising have predominantly been developed for and researched with static media stimuli (for an overview of eye-tracking research specifically; Wedel and Pieters, 2017). Extrapolating theory from static to dynamic media, as a basis for designing effective video ads for example, is problematic because the stimuli are more complex and the viewing experience differs. For example, the *viewer* decides how long to look at a print ad, where all objects are constantly present, whereas the *advertiser* decides how long an object is present to be looked at in a video ad scene by scene. This research investigates whether explaining viewer attention to video advertising content requires new theory specifically devised for dynamic stimuli, rather than theory originally developed for attention to static stimuli.

Traditional theory, such as feature integration theory (Treisman and Gelade, 1980), suggests that scenes (e.g. static photographs) are encoded using saliency variables, such as size, colour and brightness. The more salient an object is, by being bigger, more colourful or brighter than the rest of the scene, the more likely it is to attract attention. Prior research has argued that print advertising, therefore, needs to use salient elements to attract attention (Lohse, 1997; Pieters and Wedel, 2004). This traditional theory (hereon referred to as “saliency theory”) now includes measures of visual complexity, or clutter, that considers other objects competing for attention (Pieters *et al.*, 2010; Rosenholtz *et al.*, 2007). Saliency theory has also been adapted to measure attention to dynamic video by adding motion as a saliency dimension (Itti, 2004; Wolfe and Horowitz, 2004). Across contexts, the saliency theory considers these dimensions (i.e. size, colour, brightness, clutter and motion) as inherent characteristics of the environment or stimulus that will draw attention.

In contrast, newer Dynamic Human-Centred Communication Systems Theory (hereon referred to as “dynamic attention theory”), considers attention as a process of interaction between a human and the environment (Lang, 2014). According to the dynamic attention theory, humans moving through an environment will conserve energy by using elements in that environment as stores of external memory (Lang and Bailey, 2015). For example, an eye-tracking study showed that humans are cognitive misers, preferring to acquire information from the outside world using a low-effort looking strategy, rather than a high-effort memorisation strategy (Ballard *et al.*, 1997). Humans similarly treat dynamic media content as if it was a real environment (Reeves and Nass, 1996). For example, evolutionary and cultural adaptation lead us to expect that in a video of a room, the furniture and paintings will be the unchanging elements. This expectation can be exploited by change-blindness videos (Do The Test, 2008). Attention can then focus on more changeable elements in the environment, such as the people in the room.

Because the dynamic attention theory considers attention as a process of interaction between a human and the environment, this new theory predicts that dimensions of attention will interact (Lang and Bailey, 2015), whereas the salience theory assumes these dimensions are additive (Itti and Koch, 2001). One experiment, which used recognition memory to measure visual processing of media messages (including ads, public service announcements, news, sitcoms and reality programs), found results more consistent with dynamic attention theory than with salience theory (Lang and Bailey, 2015). For example, central, stable and relevant objects, which should be the most memorable because salience theory suggests visual attention will return to these objects again and again (Yegiyan and Lang, 2010), ranked the worst for recognition (16th of 16). The authors of that study argued that humans have evolved to automatically note changes in the environment and automatically perceive (encode) objects that are fleeting and peripheral (Lang and Bailey, 2015). That is, fleeting and peripheral objects must be internally memorised, which requires the viewer to attend to (and encode) them as they occur; otherwise, they will be forgotten. Further, whether an object attracts attention depends on the combined effect of its levels on four dimensions of attention, which can work in opposite directions. For example, the theory predicts that an object which is stable and relevant is more likely to be noticed if it is in the periphery of the screen (Lang and Bailey, 2015). Peripheral, stable and relevant objects ranked 5th of 16 for recognition (statistically equal with the highest-ranking objects).

Eye-tracking has been suggested as an alternative approach to further investigate the predictions of dynamic attention theory (Lang and Bailey, 2015). Eye-tracking is considered useful because it is possible that eye movements do occur in the direction of objects that should receive more attention according to the dynamic attention theory, but the resulting fixations and fixation durations may not be long enough for memory encoding. Eye-tracking, therefore, provides a more sensitive measure of the implications of dynamic attention theory than recognition memory.

Practically testing theoretical predictions about attention to dynamic stimuli requires eye-tracking software that can track attention to *moving* areas of interest (AOIs) (Wedel *et al.*, 2019; Williams and Castelhan, 2019). This feature is now available in most eye-tracking software packages but, to our knowledge, has rarely been used in the advertising literature. The specific technology we adopt is intelligent bounding boxes (IBBs) from iMotions. IBBs define the boundaries of AOIs (e.g. a logo, product or face) in a similar way to how AOIs are defined for static print advertising, but the software monitors changes in the position and dimensions of objects, both moving and reshaping with AOIs across time. Because IBBs exactly conform to the dimensions of the tracked object, fixation data are not over-estimated or under-estimated, which can occur when using static AOIs larger than the moving objects they represent (Boerman *et al.*, 2015).

Adopting this newer eye-tracking software, and potentially also newer dynamic attention theory (if further support is found), could improve our understanding of visual attention to AOIs in video ads, helping advertisers to design ads that achieve better outcomes. Advertising content can be conceptually divided into creative, branding and message elements. Branding elements (e.g. brand name, logo and other identifiers) are particularly important AOIs because if the brand is not noticed (encoded), then it is unlikely the advertising exposure will nudge the viewer's future probability to buy it. Branding elements must compete with other creative and messaging elements for viewers' limited attention. Therefore, evidence for how to execute branding elements to increase the likelihood they will get attention is of great value to advertisers.

The present study investigates these different advertising elements, building on eye-tracking research with static advertising (Pieters and Wedel, 2004) by demonstrating how IBBs can be used to measure attention to AOIs in dynamic video advertising. Moreover, differences in visual attention to creative, branding and message elements, as moving rather than static AOIs, might help to explain the differential results reported by prior studies of advertising efficacy that include these elements (Armstrong *et al.*, 2016; Hartnett *et al.*, 2020; Hartnett *et al.*, 2016; Stewart and Furse, 1986).

In summary, this study aims to make two contributions. The first is a theoretical contribution, introducing and testing the relative merits of newer dynamic attention theory versus traditional salience theory. Specifically, we investigate whether the four information dimensions of dynamic attention theory (i.e. imminence, task and motivational relevance and stability) contribute additional explained variance in eye-tracking measures beyond that explained by variables derived from salience theory (i.e. size, colour, brightness and clutter) (Pieters and Wedel, 2004; Yegiyan and Lang, 2010) and whether these attention dimensions have interactive effects, rather than simple additive effects. The second contribution is a methodological one, which is demonstrating the utility of IBBs for measuring attention to visual elements in video ads, giving guidance on how to construct scenes that direct attention to where the advertiser wants it to go.

Research background and hypotheses

The purpose of this research was to test whether attention to elements in dynamic video advertising is better explained by dynamic attention theory rather than salience theory. The key difference between the two theories, as already mentioned, is that salience theory assumes attention dimensions have additive effects (Itti and Koch, 2001), whereas dynamic attention theory predicts these dimensions will interact (Lang and Bailey, 2015).

Dynamic attention theory builds on human-computer interaction research, which similarly assumes that an active observer will use stable elements of the environment as external memory stores (in human-computer interaction research, this is called “distributed cognition”) (Halverson and Hornof, 2011; Payne *et al.*, 2001; Rogers, 2004). In a classic study, even experienced switchboard operators could not remember the positions of the numbers and letters on a telephone dial, even though that information had been central, stable and relevant in their lives for many years (Morton, 1967). Instead of memorising that information, they simply looked for it whenever they needed it.

From an active observer’s perspective, the dynamic attention theory identifies four dimensions of external information that can attract attention (Lang and Bailey, 2015):

- (1) *imminence*, defined as closeness versus distance (or central versus peripheral);
- (2) *motivational relevance*, which identifies whether the object is a survival opportunity or threat, thereby triggering appetitive or aversive systems (Cacioppo *et al.*, 2012; Carretié *et al.*, 2017);
- (3) *task relevance*, whether the object is important for the specific task at hand (e.g. the brand is a relevant object when watching advertising); and
- (4) *stability*, whether the object will remain in the environment (i.e. is stable) or could disappear (i.e. is fleeting).

Imminence, relevance and stability are also dimensions of salience theory (in which they are called, e.g. centrality, goal control and motion). But the salience theory has other variables based on fundamental aspects of visual perception, common to moving and static stimuli. In the next section (“static salience variables”), we discuss the variables unique to salience

theory, before discussing the attention dimensions captured in both theories (in “dynamic attention theory and its dimensions”).

Static salience variables

Static salience variables can have an automatic, bottom-up influence on attention. Neuroscientific investigations have revealed the importance of bottom-up salience effects on visual processing (Dehaene *et al.*, 2006). The early, automatic phase of visual processing (100 to 200 ms after stimulus onset) can include attention to not only salient elements but also emotional categorisation (Pourtois *et al.*, 2013) and even semantic processing (Carretié *et al.*, 2004). For this reason, salience, emotion and meaning, identified by initial automatic visual processing, can direct later conscious attention (Nummenmaa *et al.*, 2006).

Computer science studies of visual attention to dynamic video, including ads, show that static salience variables can explain much of the variance in attention to dynamic stimuli (Buzzelli, 2020; De Abreu *et al.*, 2017; Hou *et al.*, 2017; Itti, 2004). The static salience variables in these computer models (Itti *et al.*, 1998) include the variables from the feature-integration model of vision (Treisman and Gelade, 1980), which are detected by specialised cortical cells (Wedel and Pieters, 2006). For example, colours are processed by the parvocellular system, and brightness is processed by the magnocellular system (Carretié *et al.*, 2017). From these static salience variables, we selected those likely to characterise AOIs in video ads, and which had supporting evidence from marketing studies, to include in our research hypotheses.

The (greater) *size* of AOIs has been shown to attract and sustain attention in prior static advertising research (Lohse, 1997; Pieters and Wedel, 2004); hence, we hypothesise positive effects of size on fixation and fixation duration. *Brightness* and *colour* are also important static salience variables for visual attention (Carretié *et al.*, 2017; Detenber *et al.*, 2000; Egeth and Yantis, 1997; Gorn *et al.*, 1997; Irwin *et al.*, 2000; Rayner, 1998; Ross and Kowler, 2013; Theeuwes, 1991; von Wartburg *et al.*, 2005; Wolfe and Horowitz, 2004). For example, Lohse (1997) used eye-tracking to show that attention to Yellow Pages ads was influenced by colour, controlling for size, which justified charging more for coloured listings. More recent research shows that large size and bright colours increase the likelihood that viewers will click on ads in mobile phone apps (Mattke *et al.*, 2021).

While size, brightness and colour are expected to have positive effects on visual attention, we expect negative effects on attention from the presence of clutter (Lee *et al.*, 2018; Pieters *et al.*, 2010; Rosenholtz *et al.*, 2007). In a self-paced medium like static print ads (Ha and McCann, 2008), the time spent looking at the ad increases when there are more objects to look at (Pieters *et al.*, 2010). But in a captive medium like video (Ha and McCann, 2008), looking time is limited by scene length, which means the more objects there are to look at, the less time there will be for a viewer to look at all of them. For these reasons, clutter (i.e. the number of distracting objects) should have a negative effect on fixation and fixation duration for AOIs in dynamic media (Bennett *et al.*, 2021; Rosenholtz *et al.*, 2007; Wolfe and Horowitz, 2004).

In summary, we expect that static salience variables will explain some, but not all, of the variance in visual attention to video ads, measured by fixation and fixation duration. These arguments lead to our first hypothesis:

- H1.* Static salience variables will guide visual attention (fixation and fixation duration) to dynamic video advertising. Visual attention to areas of interest will be influenced positively by (a) size, (b) brightness and (c) colour but influenced negatively by the presence of (d) clutter.

Computer science research has considered the three basic video colours (red, green and blue) (Itti *et al.*, 1998) and, later, skin tones (Borji and Itti, 2013). In this study, we use a larger colour palette than previous research has tested (Etchebehere and Fedorovskaya, 2017) to provide more specific recommendations that advertisers can readily implement. In addition to red, green and blue, we test the attention-getting effects of orange, yellow and purple, as well as colour shades (e.g. dark orange) and neutral colours (e.g. grey). Attention to each colour was compared with attention to the absence of colour (i.e. white).

Dynamic attention theory and its dimensions

Imminence. In dynamic attention theory, imminent information is defined as information in the external environment that is closer to the perceiving human (Lang and Bailey, 2015). Imminent information is more likely to be encoded than more distant information for three reasons. First, closeness increases the relevance of biologically motivating stimuli such as dangerous predators and tasty food. Second, closeness increases the relevance of any object related to the task in hand: if the task is walking, then we need to avoid near objects in our way. Finally, closeness increases (a) the number of senses that can access information about the stimulus and, therefore, (b) the amount of sensory information gathered by each sense, which together lead to stronger memories, with more retrieval pathways for recall. In line with dynamic attention theory, closer objects in the foreground (e.g. pedestrians) attract attention in videos of natural scenes (Chun *et al.*, 2011; Wang *et al.*, 2018; Wolfe and Horowitz, 2004).

Viewers treat dynamic video as if it was a real scene, but on a flat screen, the distance between the viewer and stimuli is only *perceived*. For this reason, dynamic attention theory defines centrally located information on a screen as imminent information, as opposed to more distant information at peripheral locations. In prior research, a central location was identified as the middle area of the nine areas defined by the rule of thirds (Yegiyan and Lang, 2010). The centre of the screen is the default location for video viewing (Brasel and Gips, 2008; Yang *et al.*, 2018). Computer models of eye-tracking fixations are improved by including a centre-screen bias, as photographers tend to put the main object in the centre of a picture (Judd *et al.*, 2009; Wang *et al.*, 2018), and the edges of the video screen concentrate attention centrally, unlike natural scene viewing (Tatler *et al.*, 2011). Product placements in the centre of the screen are also considered more prominent and, therefore, more likely to be effective (Russell, 2002).

Because of the difficulty of tracking dynamic AOIs up to now, prior eye-tracking studies have examined whether fixation disperses away from this default central location (D'Ydewalle *et al.*, 1998). The association between low dispersion and the popularity of movies (Barnett and Cerf, 2017) and television programs (Dmochowski *et al.*, 2014), as well as watching, rather than avoiding, television advertising (Teixeira *et al.*, 2010), suggests that viewers prefer dynamic content that skilfully directs their attention, often by placing focal objects in the centre of the screen (Loschky *et al.*, 2015; Williams and Castelhana, 2019). However, ads can sometimes feature important elements in the corners of the screen (e.g. logos) or at the bottom (e.g. disclaimers). Prior research suggests these non-central locations will struggle to attract attention. For these reasons, we proposed the following hypothesis:

- H2. Imminence will increase visual attention to areas of interest, such that most fixations and longer fixation durations will coincide with the centre of the screen.

Motivational relevance. Dynamic attention theory is based on psychology's evaluative space model (Cacioppo *et al.*, 2012), which proposes that humans have two motivational systems, a positive (appetitive) system for approach and a negative (aversive) system for avoidance,

which are independently and automatically activated by stimuli specifically relevant to each system (Carretié *et al.*, 2017). Primary motivational stimuli (threats and opportunities e.g. food, sex and danger) have biological relevance across tasks (Lang and Bailey, 2015; Wedel and Pieters, 2006). For these reasons, salience theory research has noted that attention is attracted by *animals* (Carretié *et al.*, 2017; Judd *et al.*, 2009; Valiyamattam *et al.*, 2020) and *faces*, the latter of which are believed to be among the most (if not *the* most) biologically and socially significant stimuli for humans (Cerf *et al.*, 2007; Palermo and Rhodes, 2007; Rösler *et al.*, 2019; Rubo and Gamer, 2018). Faces rapidly attract attention from cortical cells dedicated to face processing (Tsao *et al.*, 2006). Similarly, parts of faces, such as *eyes*, attract attention (Cazzato *et al.*, 2020; Judd *et al.*, 2009). For these reasons, we expected to find evidence for our third hypothesis:

- H3. Motivational relevance will increase visual attention (fixation and fixation duration) for areas for interest representing survival threats or opportunities (e.g. human faces, eyes and animals).

Task relevance. Although automatic, bottom-up influences on attention are important, salience theory has noted that attention is also influenced by goal-driven (top-down) variables such as task instructions, which can lead people to concentrate on certain aspects of the message and ignore others (Orquin and Loose, 2013; Pieters and Wedel, 2007). Task-relevant information helps the accomplishment of a specific task, and for this reason, information relevant to achieving task goals automatically attracts attention responses (or orienting responses) (Lang and Bailey, 2015). Task instructions can change the relative importance of stimuli, by changing their relevance to the task goal (Orquin and Loose, 2013; Pieters and Wedel, 2007; Rosbergen *et al.*, 1997; Van Der Lans *et al.*, 2008; Wedel and Pieters, 2006; Yarbus, 1967). In one study, when participants were instructed to memorise print ads, they paid more visual attention to the pictures, but when instructed to learn about the brand, they paid more visual attention to the words (Pieters and Wedel, 2007).

A key task when watching video media, broadly speaking (e.g. clips, programmes or movies), is to follow the story; hence, viewers make orienting responses to stimuli that help to understand the story (e.g. events, people, actions and expectations) and encode these stimuli into memory (Lang and Bailey, 2015). When watching advertising, however, consumers have at least two tasks. The first task is understanding the story the advertiser is telling about the brand, which encourages orienting responses to structural elements, such as pictures (Pieters and Wedel, 2004) or superimposed graphics and explanatory text (Stewart and Furse, 1986). In static print ads, headline text is more likely to gain attention than body text (Pieters and Wedel, 2007; Rosbergen *et al.*, 1997). Large text (also known as “supers”) in video ads should attract attention because it is akin to a headline (Ross and Kowler, 2013; Rumpf *et al.*, 2020; Stewart and Furse, 1986). The second task consumers have when viewing ads is to assess the relevance of the advertised product or brand (Myers *et al.*, 2020; Orquin and Loose, 2013; Pieters and Wedel, 2004; Pieters and Wedel, 2007; Stewart and Furse, 1986). Eye-tracking studies of video advertising suggest that visual branding presence (either name, logo, typeface, trademark or pack shot) may have a positive effect on attracting attention (fixation) but a potential negative effect on retaining attention (fixation duration), with prolonged visual branding presence associated with increased advertising avoidance (Teixeira *et al.*, 2010). Overall, however, we expect visual branding will be relevant to viewing advertising and so will attract visual attention. These arguments lead to our fourth hypothesis:

H4. Task relevance will have a positive effect on visual attention to areas of interest, increasing fixation and fixation duration.

Stability. Dynamic attention theory conceives of humans as cognitive misers who minimise the amount of energy spent on memorising aspects of the external world (Lang, 2014; Lang and Bailey, 2015). Stable (or static) stimuli that are likely to persist in the environment do not need to be memorised; they can be used as a form of external memory (Hollan *et al.*, 2000). Visual attention and memory encoding can be focused on fleeting stimuli that need to be internally memorised to be able to access information about them later.

The need to memorise fleeting or temporary stimuli explains why movement attracts attention (Abrams and Christ, 2003; Egeth and Yantis, 1997; Smith and Abrams, 2018; Wolfe and Horowitz, 2004), except when the movement itself is stable (continuous) and so does not need to be memorised (Abrams and Christ, 2003; Folk *et al.*, 1994; Hillstrom and Yantis, 1994; Van der Burg *et al.*, 2019). We expect, therefore, that fleeting (moving) stimuli will attract greater attention, as has been shown by previous research on dynamic video using the salience theory (Dayan *et al.*, 2018; Detenber *et al.*, 1998; Reeves *et al.*, 1985; Simons *et al.*, 1999; Simons *et al.*, 2003; Yegiyani and Lang, 2010; Yegiyani and Yonelinas, 2011).

Computer models of attention to dynamic video have found that movement was more important than static salience variables for predicting fixations (Itti, 2004; Rosenholtz, 1999; Wang *et al.*, 2018). These psychological and computer science studies suggest a positive effect of movement on attracting attention, which is measured in our study by fixation. Notably, these studies provide no direct evidence about the ability of movement to retain attention, measured by fixation duration. An upper limit to fixation duration is provided by the length of time a fleeting stimulus is present on the screen. However, we expect that after controlling for the effects of the other three dimensions of dynamic attention theory (imminence, motivational relevance and task relevance), movement applied to AOIs will still explain a substantial amount of variance in fixation and fixation duration:

H5. Fleeting areas of interest will attract more attention than stable areas of interest, measured by fixation, and sustain more attention, measured by fixation duration.

Certain stimuli related to the task of processing video advertising are inherently dynamic, so we include them in this block of fleeting (moving) stimuli. A character interacting with the product to demonstrate its use (e.g. driving a car or eating a chocolate bar) very likely attracts and retains attention, which would explain why this creative device increases recall and purchase intent (Stewart and Furse, 1986). More generally, understanding the storyline in dynamic video requires paying attention to what the characters are saying, so speaking characters (or “talking heads”) should also attract attention, which can have higher-order advertising effects, such as increasing sales effectiveness (Hartnett *et al.*, 2016; Lang, 1995).

Interactions between dynamic attention theory dimensions. The novel predictions of dynamic attention theory are associated with its predictions for the interactions between its four dimensions. In contrast, salience theory, as implemented in computer models of vision (Itti, 2004), treats its variables as additive main effects (Itti and Koch, 2001). Salience theory suggests that the combination of greater salience (i.e. size, colour, brightness and low clutter), imminence and relevance should produce the most visual attention, and dynamic variables will add further attention. Instead, what Lang and Bailey (2015) proposed and found was a counterintuitive interaction effect, associated with motivationally relevant objects. Ordinarily, imminent (central) information should attract more attention than non-imminent (peripheral) information. For this reason, photographers and directors frame their shots so that story-relevant (i.e. task relevant) information is shown centrally (Williams and

Castelhana, 2019; Yegiyan and Lang, 2010). However, for motivationally and task-relevant objects, imminence and stability can suppress attention because of the relatively enduring presence and relevance of these objects. As previously discussed, whenever viewers need to remember information about stable objects, they can just look at them (Hollan *et al.*, 2000; Morton, 1967; Song *et al.*, 2019). This is useful for human survival, as it frees attention to concentrate on fleeting objects and relevant objects located peripherally (Carretié *et al.*, 2020; Lang and Bailey, 2015). These arguments lead to the following interaction hypothesis predictions:

- H6.* There will be an imminence (central vs peripheral) \times relevance (motivational or task relevant vs not) \times stability (fleeting vs stable) interaction such that: (a) for fleeting information, imminence will attract more visual attention (fixation and fixation duration) regardless of relevance; and (b) for stable information, imminence will reduce visual attention (fixation and fixation duration) if this information is motivationally or task relevant.

Materials and methods

Overview

A sample of US consumers [$N = 37$, 26 females ($M = 39.4$ years, $SD = 8.64$) and 11 males ($M = 36.9$ years, $SD = 6.52$)] watched 34 unfamiliar Australian video ads in a random order, embedded in Australian program content, on a computer screen. The screen was part of a Tobii T60 eye-tracker, which measured where participants looked (fixated) on the computer screen and for how long (fixation duration). The video ads were randomly sampled to represent the range of ads a typical consumer might see and so were for a mix of brands and product categories. The cover story for the participants was that they were watching an Australian comedy show, which had been recorded in Australia with advertising included. Randomising the order of ads, to control for potential fatigue effects, was controlled by randomly assigning participants to one of seven pre-set variations.

Potential AOIs were identified in each scene in each ad, and the number of scenes ranged from 11 to 28 across ads. The maximum number of potential AOIs in each scene was capped at seven, which is the capacity of working memory (Miller, 1956). Using the eye-tracking software, IBBs were drawn around these AOIs, and these IBBs automatically adjusted to fit around the AOIs if they moved or changed size over the duration of the scene or across scenes. The eye-tracking data from each participant could then be coded as fixation = 1 (otherwise 0) if a fixation occurred inside one of these AOIs. Whenever fixation = 1, fixation duration in milliseconds was also recorded. The AOIs were coded for the presence of static salience variables (size, colour, brightness and clutter) or dynamic attention theory variables (imminence, motivational or task relevance and stability). Regression was used to estimate the effects of these variables on fixation and fixation duration (see Appendix 1 for more information about the methods and analysis used).

Results

Static salience variables

H1 predicted that static salience variables will guide visual attention (fixation and fixation duration) to video advertising. That is, visual attention to AOIs will be influenced positively by increases in (a) size, (b) brightness and (c) colour but negatively by (d) clutter.

This general hypothesis was supported by the amount of variance in visual attention explained by these variables. Static salience variables contributed 90% of total explained

variance in fixation and 57% of explained variance in fixation duration. Not every sub-hypothesis of *H1* was supported, however.

Size had positive effects on fixation ($b = 5.40$, $SE = 0.06$, $\exp(b) = 221.21$, Wald $\chi^2(1) = 7,816.56$ and $p < 0.001$) and fixation duration ($b = 0.62$, $SE = 0.02$, $\beta = 0.25$, $t = 39.02$ and $p < 0.001$), supporting *H1a*. Brightness had a significant effect on fixation duration only, and it was a negative effect ($b = -0.12$, $SE = 0.02$, $\beta = -0.03$, $t = -6.90$ and $p < 0.001$), contrary to *H1b*. Regarding colour, results gave mixed support for *H1c*. The colour block added only 1.1% to total explained variance in fixation and 1.5% in fixation duration. All colours were expected to have positive effects on visual attention, compared with no colour (white) and controlling for brightness. Yellow had a large positive effect on fixation ($b = 5.01$, $SE = 0.58$, $\exp(b) = 149.65$, Wald $\chi^2(1) = 73.29$ and $p < 0.001$), but other colours (e.g. orange, red and dark blue) had negative effects or no effect (blue). Clutter, which affected 62% of AOIs, had its expected negative effect on fixation ($b = -0.36$, $SE = 0.02$, $\exp(b) = 0.70$, Wald $\chi^2(1) = 451.33$ and $p < 0.001$), supporting *H1d*, but its effect on fixation duration was not significant (see [Appendix 2](#) for more detailed results for *H1* and the other hypotheses).

Dynamic attention theory and its dimensions

Imminence. *H2* predicted that imminence (central screen location) will increase visual attention to AOIs, such that most fixations and longer durations will coincide with the centre of the screen. The imminence block contributed 6% of total explained variance in fixation, but only 0.2% of explained variance in fixation duration. As predicted by *H2*, a central location was significantly and positively associated with fixation ($b = 0.41$, $SE = 0.03$, $\exp(b) = 1.51$, Wald $\chi^2(1) = 243.64$ and $p < 0.001$), but contrary to *H2*, central location was significantly negative for fixation duration ($b = -0.20$, $SE = 0.01$, $\beta = -0.11$, $t = -16.38$ and $p < 0.001$). *H2* is, therefore, only partially supported.

Motivational relevance. *H3* predicted that motivational relevance will increase fixation and fixation duration for certain AOIs representing survival threats or opportunities (e.g. human faces, eyes and animals). The block of motivationally relevant variables contributed only 0.2% to total explained variance in fixation and 0.3% to fixation duration. Contrary to *H3*, motivational relevance was associated with significant negative effects on fixation ($b = -0.21$, $SE = 0.03$, $\exp(b) = 0.81$, Wald $\chi^2(1) = 62.10$ and $p < 0.001$) and fixation duration ($b = -0.24$, $SE = 0.02$, $\beta = -0.12$, $t = -16.25$ and $p < 0.001$). Hence, these results do not provide support for *H3*.

Task relevance. *H4* predicted that task relevance will have a positive effect on visual attention to AOIs (e.g. graphics, text and branding elements), increasing fixation and fixation duration. Like motivational relevance, the block of task relevant variables did not contribute much to explained variance, only 0.2% for fixation and 0.5% for fixation duration. Also, contrary to *H4*, task relevance had a significant negative effect on fixation ($b = -0.25$, $SE = 0.03$, $\exp(b) = 0.78$, Wald $\chi^2(1) = 73.59$ and $p < 0.001$) and fixation duration ($b = -0.36$, $SE = 0.02$, $\beta = -0.16$, $t = -24.28$ and $p < 0.001$). Therefore, the results do not provide support for *H4*.

Stability. *H5* predicted that fleeting AOIs will attract and sustain more attention than stable AOIs, measured by fixation and fixation duration, respectively. The stability variables block contributed only 0.3% of total explained variance in fixation but 10% of total explained fixation duration. Contrary to *H5*, the effect of fleeting versus stable AOIs was significantly negative for both fixation ($b = -0.49$, $SE = 0.03$, $\exp(b) = 0.62$, Wald $\chi^2(1) = 277.25$ and $p < 0.001$) and fixation duration ($b = -0.48$, $SE = 0.01$, $\beta = -0.21$, $t = -47.32$ and $p < 0.001$). Consequently, *H5* was not supported.

Interactions between dynamic attention theory dimensions. *H6* predicted a significant interaction between the dynamic attention theory dimensions of imminence (central vs peripheral), relevance (motivational or task relevant vs not) and stability (fleeting vs stable). The interaction variables block contributed 1% of total explained variance in fixation and 4% of explained variance in fixation duration. *H6* was supported by the significant effects of the three-way interaction between imminence, *motivational* relevance and stability on fixation ($b = 0.28$, $SE = 0.04$, $\exp(b) = 1.33$, Wald $\chi^2(1) = 64.36$ and $p < 0.001$) and fixation duration ($b = 0.30$, $SE = 0.02$, $\beta = 0.13$, $t = 16.91$ and $p < 0.001$). There were also significant effects of the interaction between imminence, *task* relevance and stability on fixation ($b = 0.66$, $SE = 0.04$, $\exp(b) = 1.94$, Wald $\chi^2(1) = 248.58$ and $p < 0.001$) and fixation duration ($b = 0.47$, $SE = 0.02$, $\beta = 0.17$, $t = 25.05$ and $p < 0.001$). Replicating [Lang and Bailey \(2015\)](#), [Figure 1](#) depicts the two-way interactions between imminence and relevance (motivational and task relevance) for fleeting versus stable AOIs for fixation duration (results for fixation were similar).

H6a predicted that for fleeting information, imminence will attract more visual attention regardless of relevance. The left-hand charts in [Figure 1](#) show that for fleeting information, imminence increased the effects of motivational and task relevance on fixation duration. In the top-left chart, for fleeting motivationally relevant information, fixation duration was longer when that information was central [460 ms, 95% CI (449, 471)] rather than peripheral [414 ms, 95% CI (401, 428)]. The bottom-left chart shows that for fleeting task-relevant information, fixation duration was also longer for central versus peripheral positions [485 ms, 95% CI (473, 498) vs 370 ms, 95% CI (356, 384)]. However, the effect of imminence did not occur *regardless* of relevance. If task or motivational relevance was absent, then a central location, versus a peripheral location, reduced fixation duration [central 432 ms, 95% CI (425, 439) vs peripheral 528 ms, 95% CI (520, 537)]. Consequently, *H6a* was only partially supported, as imminence did not improve attention to fleeting objects regardless of relevance. Specifically, the positive effect of imminence was associated *only* with relevant fleeting objects.

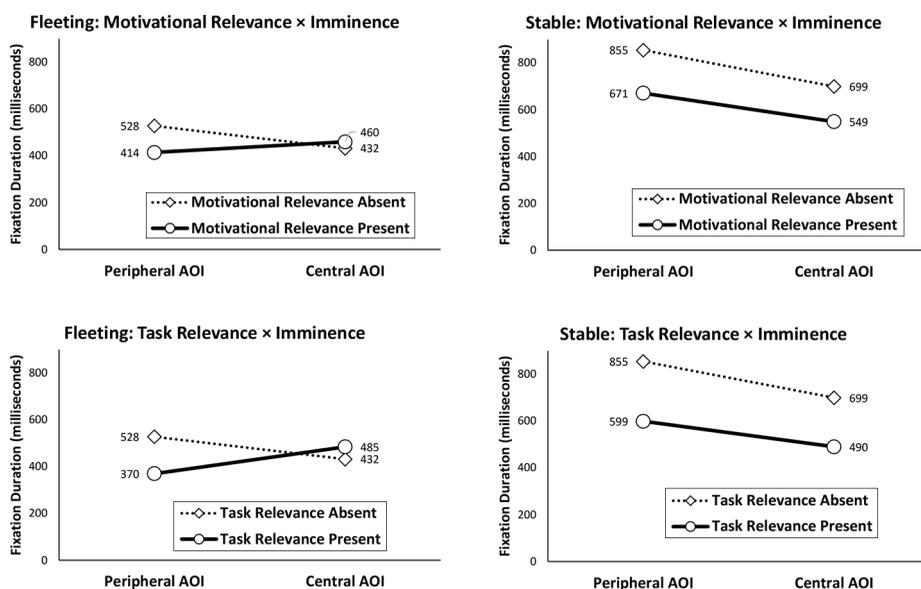


Figure 1. Three-way interaction effects on fixation duration between stability (fleeting vs stable), imminence (central vs peripheral) and motivational relevance (top row) versus task relevance (bottom row)

H6b predicted that for stable information, imminence will reduce visual attention if information is motivationally or task relevant. The top-right chart in [Figure 1](#) shows that for stable information, imminence reduced the effect of motivational relevance on fixation duration [central 549 ms, 95% CI (470, 642) vs peripheral 671 ms, 95% CI (584, 770)]. The bottom-right chart shows similar results for task relevance, with imminence reducing fixation duration [central 490 ms (95% CI = 449, 534) vs peripheral 599 ms (95% CI = 551, 651)]. Further to this, at central locations, fixation duration was significantly lower when motivational relevance was present [549 ms, 95% CI (470, 642) vs absent 699 ms, 95% CI (674, 726)], and when task relevance was present [490 ms, 95% CI (449, 534) vs absent 699 ms, 95% CI (643, 762)]. These results supported *H6b*.

Discussion

This research tested the proposition that dynamic advertising content needs a new dynamic attention theory to explain where consumers will direct their visual attention. Our results compared the conflicting predictions of salience theory ([Christianson et al., 1991](#); [Itti et al., 1998](#); [Russell, 2002](#); [Treisman and Gelade, 1980](#); [Yegiyen and Lang, 2010](#)) versus the novel predictions of dynamic attention theory ([Lang, 2014](#); [Lang and Bailey, 2015](#)). Consistent with previous research using search tasks ([Folk et al., 1994](#)) or natural scenes ([Carmi and Itti, 2006](#)), static salience variables attracted and retained attention, even when motion was present. However, dynamic attention theory uniquely predicted that a central location would not *always* increase attention. These results have theoretical implications for future attention research and practical implications for designing and testing video ads.

Implications for theory

This study was a first step in a new line of research, investigating how dynamic attention theory ([Lang and Bailey, 2015](#)) could improve visual attention to dynamic marketing stimuli, such as video ads. We found evidence consistent with dynamic attention theory rather than salience theory, suggesting that dynamic attention theory provides a more useful framework for understanding and predicting visual attention to video ads. Although dynamic attention theory was developed by media psychology researchers and draws on prior research in the field of human computer interaction, our findings suggest this new theory will make an important contribution to the marketing literature in the future.

Dynamic attention theory proposes four dimensions of external information:

- (1) imminence (central vs peripheral location);
- (2) motivational relevance (objects related to survival threats or opportunities);
- (3) task relevance (in our experiment, relevance to understanding what the advertising was about); and
- (4) stability (fleeting vs stable).

As shown in [Table 1](#), salience theory also proposes these four dimensions (with different names, e.g. “centre bias”, “faces”, “goal driven” and “motion”) but assumes they will have additive main effects (*H2* to *H5*) on gaining and sustaining attention (fixation and fixation duration). However, only dynamic attention theory predicts that these dimensions will interact (*H6*), and one of the consequences of a significant interaction effect may be that the main effects of these dimensions are not statistically significant or significant in the opposite direction to the one predicted. Our results, in which we found significant interactions between the four dimensions, affecting both fixation and fixation duration, support dynamic

Hypothesis	Theory	Supported
<i>H1</i> . Static salience variables will guide visual attention (fixation and fixation duration) to dynamic video advertising. Visual attention to AOs will be influenced positively by (a) size, (b) brightness and (c) colour but influenced negatively by the presence of (d) clutter	Salience	Partially, we found support for <i>H1a</i> , no support for <i>H1b</i> and mixed support for <i>H1c</i> and <i>H1d</i>
<i>H2</i> . Imminence will increase visual attention to AOs, such that most fixations and longer fixation durations will coincide with the centre of the screen	Salience (main effect)	Partially, we found support for fixation but not for fixation duration
<i>H3</i> . Motivational relevance will increase visual attention (fixation and fixation duration) for AOs representing survival threats or opportunities	Salience (main effect)	No
<i>H4</i> . Task relevance will have a positive effect on visual attention to AOs, increasing fixation and fixation duration	Salience (main effect)	No
<i>H5</i> . Fleeting AOs will attract more attention than stable AOs, measured by fixation, and sustain more attention, measured by fixation duration	Salience (main effect)	No
<i>H6</i> . There will be an imminence (central vs peripheral) \times relevance (low relevance/ motivationally relevant/task relevant) \times stability (fleeting vs stable) interaction such that: (a) for fleeting information, imminence will attract more visual attention (fixation and fixation duration) regardless of relevance; and (b) for stable information, imminence will reduce visual attention (fixation and fixation duration) if this information is motivationally or task relevant	Dynamic attention	Yes

Table 1.
Summary of results for hypothesis tests

attention theory rather than salience theory, even though static salience variables explained most of the variance in fixation and fixation duration in our data.

Further evidence supporting dynamic attention theory, as opposed to salience theory, comes from the pattern of the significant interaction effects, which reproduced results previously reported by [Lang and Bailey \(2015\)](#) for general video content (including but not limited to ads and public service announcements), now extended to a larger sample of video ads exclusively. [Lang and Bailey \(2015\)](#) predicted and found that stable, motivationally relevant information at the periphery of the screen (not imminent) will be remembered better than when it is in the centre (imminent). The reason for this is that the human body has evolved over time to use its environment as a store of external memory ([Hollan et al., 2000](#); [Morton, 1967](#); [Song et al., 2019](#)), conserving energy for memorising potentially disappearing (fleeting) information that may be needed later. Imminent, stable and relevant information is likely to be imminent, stable and relevant in the future, so it does not need effortful attention and memorising. [Lang and Bailey \(2015\)](#) reported a significant interaction only for motivational relevance, but we extend their results to report a significant interaction for information relevant to the task of processing video ads; imminence also decreased attention to stable task relevant information.

The reason we found significant interactions with both kinds of relevance may be that we measured processes of attention, using eye-tracking, whereas [Lang and Bailey \(2015\)](#) measured outcomes of attention, using a recognition task. Our results confirm [Lang and Bailey's \(2015\)](#) speculation that eye-tracking would be a more sensitive attention measure than recognition, because fleeting peripheral objects might attract visual attention but not remain on-screen long enough to be reliably encoded into memory. [Lang and Bailey \(2015\)](#) also speculated that video content's centre bias ([Loschky et al., 2015](#)) might even prevent eye-tracking from detecting the attention to peripheral objects predicted by dynamic attention theory. Our results show, however, that the attention effects predicted by dynamic attention theory can override the strong influence of centre bias. If future research was to investigate amateur video posts (e.g. consumer generated content), where there is less photographer control over where viewers look, then the effects predicted by dynamic attention theory may be even stronger ([Dorr et al., 2010](#)).

Future eye-tracking research could investigate the reasons why the main effect hypotheses of salience and dynamic attention theory were not always supported, specifically, task relevance (*H3*), motivational relevance (*H4*) and stability (*H5*). For example, in advertising, which has a commercial imperative, visual branding is task relevant for understanding and recognising ads as ads, but branding needs to be used in shorter pulses, with frequency ([Romaniuk, 2009](#)), to reduce the chance of a negative avoidance reaction from the viewer ([Teixeira et al., 2010](#)). Task relevance may have its hypothesised positive effect on attracting visual attention to other types of advertising objects that we did not code for, for example, unbranded story elements ([Lang and Bailey, 2015](#)), which are essential to video drama but not always found in advertising ([Kim et al., 2017](#)).

We coded for the presence of elements related to dynamic attention theory's dimensions in a sample of real video ads. But future research should manipulate the presence, absence and effect size of these elements to provide causal evidence of their effects. For example, by manipulating the extent and duration of motion, future research could identify the boundary conditions for when fleeting objects attract attention, versus evading attention, which was their main (negative) effect in our results. Just as animals and faces can go unnoticed because of attentional limitations ([Cohen et al., 2012](#)), physical limitations on eye movement may mean that moving objects need to be present on-screen for a minimum time to attract fixation ([Hinde et al., 2017](#); [Pieters et al., 2010](#)) and significant fixation duration ([Abrams and Christ, 2003](#); [Lang and Bailey, 2015](#); [Posner and Cohen, 1984](#)). Dynamic attention theory further suggests a potential inverse-U effect of time on-screen for moving objects, such that past a certain length of time on-screen, the moving object is no longer novel and attention moves elsewhere ([Lang and Bailey, 2015](#); [Van der Burg et al., 2019](#)).

Implications for advertisers

Considering the growth of dynamic media advertising opportunities, including online video and digital billboards, an important implication of these findings for advertisers is that they can conveniently use eye-tracking to measure consumers' attention to moving elements in video advertising in dynamic media. The present study demonstrates the ability of newer eye-tracking software (IBBs) to measure visual attention toward moving AOs present in video ads ([Wedel et al., 2019](#)). Previously, when eye-tracking has been applied to video stimuli, the objects it could track had to be largely static, such as disclosure messages at the bottom of the screen ([Boerman et al., 2015](#)) or billboards in sports video content ([Breuer and Rumpf, 2012](#); [Rumpf et al., 2020](#)). Eye-tracking applied to video advertising has used heat maps to provide qualitative evidence about which areas of the screen attract attention across

space and time. Now, with IBBs, advertisers can use quantitative data (fixation and fixation duration) to compare attention to different AOIs. We showed that with IBB software, which is commonly available (Wedel *et al.*, 2019) but not commonly applied in the advertising literature, advertisers can track any moving object in video ads, spanning branding and creative elements.

Advertisers frequently take advantage of the ability of video to show moving objects, such as people using or enjoying the product (Hartnett *et al.*, 2016; Stewart and Furse, 1986). Newer IBBs software allows advertisers to test whether visual attention to moving elements (including creative, branding and message) explains why their video ads are sales successful or not (Bellman *et al.*, 2019). For example, in our sample of ads, attention to fleeting (vs stable) AOIs explained more variance in fixation duration than attention to task relevant items (i.e. visual branding). This suggests that movement distracts attention from branding and that moving branding is more likely to get attention. The capability to track attention to moving objects is a timely development likely to grow in popularity, as video advertising is increasing in functionality and popularity on smartphones and digital billboards, as well as on television and computer screens [e.g. with on demand broadcaster video and over-the-top (OTT) providers, such as Amazon Prime Video and Hulu].

Our significant interaction results, which favour dynamic attention theory (Lang, 2014; Lang and Bailey, 2015), rather than salience theory, suggest that advertisers should use this newer theory to guide the design of video ads. For example, dynamic attention theory suggests that the centre position may not be the best position to place certain types of content. According to dynamic attention theory, task relevant content (e.g. branding) is more likely to gain attention at peripheral locations on the screen.

Another implication for advertisers from this study is the use of image manipulation software to quantify the colours in AOIs. Until now, most research concerning the impact of colour on visual attention has used simple colours (e.g. red vs blue: Anllo-Vento *et al.*, 1998) or the average RGB values (i.e. red, green and blue) for the pixels at a screen location (Itti *et al.*, 1998). These RGB values can define over 16 million colours. Our colour-coding scheme uses a readily understood reduced palette of colours, allowing measurement and testing, of specific colours. Our results suggest that yellow has exceptional attention-getting properties (Etchebehere and Fedorovskaya, 2017). However, this may have been because of yellow's rarity on-screen in our sample of ads, so this result may not be generalisable.

Limitations and directions for future research

This study had several limitations that suggest the need for replication and future research. The sample of consumers was small, reduced by calibration failure, resulting in more female than male participants. The results of this study, therefore, have limited external validity. However, the final sample of 37 participants is comparable with other studies using a within-participants, repeated measures design (Guerreiro *et al.*, 2015: $N = 41$; Ravaja *et al.*, 2013: $N = 33$), and the fact that it replicated a large student-sample study (Lang and Bailey, 2015) suggests that its main theory-testing results have internal validity and will likely replicate in future studies.

Another limitation is the small, random sample of ads used, which were all from one country and were all 30-second television ads. While the results of this study can apply to other media, such as video ads on computers, smartphones and digital billboards, advertisers will need to be even more efficient to capture attention in these media where video ads are typically much shorter (Campbell and Pearson, 2021). Future studies, using different samples of ads, will potentially yield different results. A further limitation of using real ads, rather than constructed stimuli, is that many potential AOIs were never fixated on,

creating a data set with many zeroes (“zero inflation”). Some previous eye-tracking studies have deleted zero fixation AOIs (Borji and Itti, 2013), but we retained all AOIs (observed in four or more ads) because advertisers are perhaps just as interested in what does not attract attention as what does. By using fixation and fixation duration, we combined the strengths of studies examining what does and does not attract attention (fixation), and studies of what attracts more versus less attention when zero attention is ignored (fixation duration). Zero inflation will be higher in future studies that use more than the maximum of seven potential AOIs we identified in each scene. Another limitation of using real ads, as touched on previously, is that the independent variables were measured, not manipulated, and so our results are correlational, not causal. Future eye-tracking studies should use controlled experiments manipulating and counterbalancing the dimensions of dynamic attention theory.

We measured only fixation and fixation duration, which are “density and duration” measures (Jacob and Karn, 2003), as opposed to “sequence and transition” measures, such as first fixation, time until first fixation and order of fixations (Wedel *et al.*, 2019). Measuring these variables in future research would increase our understanding of how viewers “read” a scene in a video advertisement (Land and McLeod, 2000). Future research should also look to report the number of fixations inside AOIs, which has demonstrated a stronger relationship with memory compared with fixation duration (Orquin and Holmqvist, 2018).

It is difficult to move the eyes quickly from the default centre of the screen to fixate on moving objects in the periphery of the screen (Cohen *et al.*, 2012). The likelihood of fixation and longer fixation duration times (see Figure 1 and Appendix 2) were both improved by a longer (more stable) time on-screen (Lang and Bailey, 2015), where time on-screen is limited by scene-length in videos. As it is difficult to fixate (i.e. rest the eyes) on a moving object, future research might analyse eye movements (saccades) rather than fixations (Rayner *et al.*, 2009). The effect of moving objects on attracting visual attention may be that people make saccades in their direction but rarely land the eye long enough to make a fixation (Lang and Bailey, 2015).

We also measured only the process of visual attention, using eye-tracking, rather than the outcomes of visual attention on downstream variables like memory, attitudes and behaviour. Future research should include these additional measures, as eye fixations do not necessarily mean that information was encoded in memory (Orquin and Holmqvist, 2018). Eye-tracking is an expensive method, which limits sample size, so future researchers might consider using cheaper self-report measures, such as recognition and recall, and other behavioural observation measures, such as heart rate (Rubo and Gamer, 2018), to test the implications of dynamic attention theory. These methods can measure attention outside of the laboratory, in distracted home environments where people may not pay much attention to video advertising (Jayasinghe and Ritson, 2013). Memory and heart rate can also measure the effects of non-visual branding, such as audio brand mentions and jingles (Simmonds *et al.*, 2020), which eye-tracking is blind to.

Conclusion

The goal of this study was to demonstrate how eye-movement research on attention to video advertising could benefit from two relatively recent innovations. The first innovation was the application of dynamic attention theory to video ads (Lang, 2014; Lang and Bailey, 2015). Dynamic attention theory predicts that the dimensions of attention will interact, whereas salience theory predicts that these dimensions have additive effects (Itti and Koch, 2001). With a random sample of video ads and a non-student sample of consumers, we used eye-

tracking to examine visual attention and found evidence of interaction effects, supporting dynamic attention theory rather than salience theory. The second innovation was to adopt a newer eye-tracking software feature (IBBs) for measuring attention to “doubly dynamic” stimuli, when objects *and* the viewer’s eyes are both moving (Wedel *et al.*, 2019). We hope that the successful demonstration of these two innovations will influence future researchers to use them, to further develop our understanding of how to improve visual attention to dynamic content in video advertising.

References

- Abrams, R.A. and Christ, S.E. (2003), “Motion onset captures attention”, *Psychological Science*, Vol. 14 No. 5, pp. 427-432.
- Anllo-Vento, L., Luck, S.J. and Hillyard, S.A. (1998), “Spatio-temporal dynamics of attention to color: evidence from human electrophysiology”, *Human Brain Mapping*, Vol. 6 No. 4, pp. 216-238.
- Armstrong, J.S., Du, R., Green, K.C. and Graefe, A. (2016), “Predictive validity of evidence-based persuasion principles: an application of the index method”, *European Journal of Marketing*, Vol. 50 Nos 1/2, pp. 276-293.
- Ballard, D., Hayhoe, M., Pook, P. and Rao, R. (1997), “Deictic codes for the embodiment of cognition”, *Behavioral and Brain Sciences*, Vol. 20 No. 4, pp. 723-767.
- Barnett, S.B. and Cerf, M. (2017), “A ticket for your thoughts: method for predicting content recall and sales using neural similarity of moviegoers”, *Journal of Consumer Research*, Vol. 44 No. 1, pp. 160-181.
- Bellman, S., Nencycz-Thiel, M., Kennedy, R., Hartnett, N. and Varan, D. (2019), “Best measures of attention to creative tactics in TV advertising: when do attention-getting devices capture or reduce attention?”, *Journal of Advertising Research*, Vol. 59 No. 3, pp. 295-311.
- Bennett, C.R., Bex, P.J. and Merabet, L.B. (2021), “Assessing visual search performance using a novel dynamic naturalistic scene”, *Journal of Vision*, Vol. 21 No. 1, pp. 1-14.
- Boerman, S.C., Van Reijmersdal, E.A. and Neijens, P.C. (2015), “Using eye tracking to understand the effects of brand placement disclosure types in television programs”, *Journal of Advertising*, Vol. 44 No. 3, pp. 196-207.
- Borji, A. and Itti, L. (2013), “State-of-the-art in visual attention modeling”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35 No. 1, pp. 185-207.
- Brasel, S.A. and Gips, J. (2008), “Points of view: where do we look when we watch TV?”, *Perception*, Vol. 37 No. 12, pp. 1890-1894.
- Breuer, C. and Rumpf, C. (2012), “The viewer’s reception and processing of sponsorship information in sport telecasts”, *Journal of Sport Management*, Vol. 26 No. 6, pp. 521-531.
- Buzzelli, M. (2020), “Recent advances in saliency estimation for omnidirectional images, image groups, and video sequences”, *Applied Sciences*, Vol. 10 No. 15, pp. 5143-5174.
- Cacioppo, J.T., Berntson, G.G., Norris, C.J. and Gollan, J.K. (2012), “The evaluative space model”, in van Lange, P.A.M., Kruglanski, A.W. and Higgins, E.T. (Eds), *Handbook of Theories of Social Psychology*, Sage Publications, London, pp. 50-73.
- Campbell, C. and Pearson, E. (2021), “Strategies for more effective six-second video advertisements: making the most of 144 frames”, *Journal of Advertising Research*, Vol. 61 No. 3, pp. 260-275.
- Carmi, R. and Itti, L. (2006), “Visual causes versus correlates of attentional selection in dynamic scenes”, *Vision Research*, Vol. 46 No. 26, pp. 4333-4345.
- Carretié, L., Hinojosa, J.A., Martín-Loeches, M., Mercado, F. and Tapia, M. (2004), “Automatic attention to emotional stimuli: neural correlates”, *Human Brain Mapping*, Vol. 22 No. 4, pp. 290-299.

- Carretié, L., Kessel, D., García-Rubio, M.J., Giménez-Fernández, T., Hoyos, S. and Hernández-Lorca, M. (2017), "Magnocellular bias in exogenous attention to biologically salient stimuli as revealed by manipulating their luminosity and color", *Journal of Cognitive Neuroscience*, Vol. 29 No. 10, pp. 1699-1711.
- Carretié, L., Méndez-Bértolo, C., Bódalo, C., Hernández-Lorca, M., Fernández-Folgueiras, U., Fondevila, S. and Giménez-Fernández, T. (2020), "Retinotopy of emotion: perception of negatively valenced stimuli presented at different spatial locations as revealed by event-related potentials", *Human Brain Mapping*, Vol. 41 No. 7, pp. 1711-1724.
- Cazzato, D., Leo, M., Distante, C. and Voos, H. (2020), "When I look into your eyes: a survey on computer vision contributions for human gaze estimation and tracking", *Sensors*, Vol. 20 No. 13, pp. 3739-3780.
- Cerf, M., Harel, J., Einhäuser, W. and Koch, C. (2007), "Predicting human gaze using low-level saliency combined with face detection", *Advances in Neural Information Processing Systems*, pp. 241-248.
- Christianson, S.-A., Loftus, E.F., Hoffman, H. and Loftus, G.R. (1991), "Eye fixations and memory for emotional events", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 17 No. 4, pp. 693-701.
- Chun, M.M., Golomb, J.D. and Turk-Browne, N.B. (2011), "A taxonomy of external and internal attention", *Annual Review of Psychology*, Vol. 62 No. 1, pp. 73-101.
- Cohen, M.A., Cavanagh, P., Chun, M.M. and Nakayama, K. (2012), "The attentional requirements of consciousness", *Trends in Cognitive Sciences*, Vol. 16 No. 8, pp. 411-417.
- Dayan, E., Barliya, A., de Gelder, B., Hendler, T., Malach, R. and Flash, T. (2018), "Motion cues modulate responses to emotion in movies", *Scientific Reports*, Vol. 8 No. 1, pp. 1-10.
- De Abreu, A., Ozcinar, C. and Smolic, A. (2017), "Look around you: saliency maps for omnidirectional images in VR applications", *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, IEEE, Piscataway, NJ, pp. 1-6.
- Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J. and Sergent, C. (2006), "Conscious, preconscious, and subliminal processing: a testable taxonomy", *Trends in Cognitive Sciences*, Vol. 10 No. 5, pp. 204-211.
- Dentsu (2021), "Global Ad spend forecasts", London, UK.
- Detenber, B.H., Simons, R.F. and Bennett, G.G. (1998), "Roll'em!: the effects of picture motion on emotional responses", *Journal of Broadcasting and Electronic Media*, Vol. 42 No. 1, pp. 113-127.
- Detenber, B.H., Simons, R.F. and Reiss, J.E. (2000), "The emotional significance of color in television presentations", *Media Psychology*, Vol. 2 No. 4, pp. 331-355.
- Do The Test (2008), "Test your awareness: Whodunnit?", Directed by Do The Test YouTube, UK.
- Dmochowski, J.P., Bezdek, M.A., Abelson, B.P., Johnson, J.S., Schumacher, E.H. and Parra, L.C. (2014), "Audience preferences are predicted by temporal reliability of neural processing", *Nature Communications*, Vol. 5 No. 1, pp. 45-67.
- Dorr, M., Martinetz, T., Gegenfurtner, K.R. and Barth, E. (2010), "Variability of eye movements when viewing dynamic natural scenes", *Journal of Vision*, Vol. 10 No. 10, pp. 28-28.
- D'Ydewalle, G., Desmet, G. and Van Rensbergen, J. (1998), "Film perception: the processing of film cuts", in Underwood, G. (Ed.), *Eye Guidance in Reading and Scene Perception*, Elsevier Science, Amsterdam, pp. 357-367.
- Egeth, H.E. and Yantis, S. (1997), "Visual attention: control, representation, and time course", *Annual Review of Psychology*, Vol. 48 No. 1, pp. 269-297.
- Etchebehere, S. and Fedorovskaya, E. (2017), "On the role of color in visual saliency", *Electronic Imaging*, Vol. 29 No. 14, pp. 58-63.

- Folk, C.L., Remington, R.W. and Wright, J.H. (1994), "The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color", *Journal of Experimental Psychology*, Vol. 20 No. 2, pp. 317-329.
- Gorn, G., Chattopadhyay, A., Yi, T. and Dahl, D. (1997), "Effects of color as an executional cue in advertising: they're in the shade", *Management Science*, Vol. 43 No. 10, pp. 1387-1400.
- Guerreiro, J., Rita, P. and Trigueiros, D. (2015), "Attention, emotions and cause-related marketing effectiveness", *European Journal of Marketing*, Vol. 49 Nos 11/12, pp. 1728-1750.
- Ha, L. and McCann, K. (2008), "An integrated model of advertising clutter in offline and online media", *International Journal of Advertising*, Vol. 27 No. 4, pp. 569-592.
- Halverson, T. and Hornof, A.J. (2011), "A computational model of 'active vision' for visual search in human-computer interaction", *Human-Computer Interaction*, Vol. 26 No. 4, pp. 285-314.
- Hartnett, N., Kennedy, R., Sharp, B. and Greenacre, L. (2016), "Creative that sells: how advertising execution affects sales", *Journal of Advertising*, Vol. 45 No. 1, pp. 102-112.
- Hartnett, N., Greenacre, L., Kennedy, R. and Sharp, B. (2020), "Extending validity testing of the persuasion principles index", *European Journal of Marketing*, Vol. 54 No. 9, pp. 2245-2255.
- Hillstrom, A.P. and Yantis, S. (1994), "Visual motion and attentional capture", *Perception and Psychophysics*, Vol. 55 No. 4, pp. 399-411.
- Hinde, S.J., Smith, T.J. and Gilchrist, I.D. (2017), "In search of oculomotor capture during film viewing: implications for the balance of top-down and bottom-up control in the saccadic system", *Vision Research*, Vol. 134, pp. 7-17.
- Hollan, J., Hutchins, E. and Kirsh, D. (2000), "Distributed cognition: toward a new foundation for human-computer interaction research", *ACM Transactions on Computer-Human Interaction*, Vol. 7 No. 2, pp. 174-196.
- Hou, Q., Cheng, M.-M., Hu, X., Borji, A., Tu, Z. and Torr, P.H.S. (2017), "Deeply supervised salient object detection with short connections", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815-828.
- Irwin, D.E., Colcombe, A.M., Kramer, A.F. and Hahn, S. (2000), "Attentional and oculomotor capture by onset, luminance and color singletons", *Vision Research*, Vol. 40 Nos 10/12, pp. 1443-1458.
- Itti, L. (2004), "Automatic foveation for video compression using a neurobiological model of visual attention", *IEEE Transactions on Image Processing*, Vol. 13 No. 10, pp. 1304-1318.
- Itti, L. and Koch, C. (2001), "Feature combination strategies for saliency-based visual attention systems", *Journal of Electronic Imaging*, Vol. 10 No. 1, pp. 161-169.
- Itti, L., Koch, C. and Niebur, E. (1998), "A model of saliency-based visual attention for rapid scene analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20 No. 11, pp. 1254-1259.
- Jacob, R.J.K. and Karn, K.S. (2003), "Eye tracking in human-computer interaction and usability research: ready to deliver the promises", in Hyönä, J., Radach, R. and Deubel, H. (Eds), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, Elsevier, North Holland, Amsterdam, pp. 573-605.
- Jayasinghe, L. and Ritson, M. (2013), "Everyday advertising context: an ethnography of advertising response in the family living room", *Journal of Consumer Research*, Vol. 40 No. 1, pp. 104-121.
- Judd, T., Ehinger, K., Durand, F. and Torralba, A. (2009), "Learning to predict where humans look", Paper presented at the 12th International Conference on Computer Vision, Kyoto, Japan (accessed October 2020).
- Kim, E., Ratneshwar, S. and Thorson, E. (2017), "Why narrative ads work: an integrated process explanation", *Journal of Advertising*, Vol. 46 No. 2, pp. 283-296.

- Land, M.F. and McLeod, P. (2000), "From eye movements to actions: how batsmen hit the ball", *Nature Neuroscience*, Vol. 3 No. 12, pp. 1340-1345.
- Lang, A. (1995), "Defining audio/video redundancy from a limited-capacity information processing perspective", *Communication Research*, Vol. 22 No. 1, pp. 86-115.
- Lang, A. (2014), "Dynamic human-centered communication systems theory", *The Information Society*, Vol. 30 No. 1, pp. 60-70.
- Lang, A. and Bailey, R.L. (2015), "Understanding information selection and encoding from a dynamic, energy saving, evolved, embodied, embedded perspective", *Human Communication Research*, Vol. 41 No. 1, pp. 1-20.
- Lee, J.E., Hur, S. and Watkins, B. (2018), "Visual communication of luxury fashion brands on social media: effects of visual complexity and brand familiarity", *Journal of Brand Management*, Vol. 25 No. 5, pp. 449-462.
- Lohse, G.L. (1997), "Consumer eye movement patterns on yellow pages advertising", *Journal of Advertising*, Vol. 26 No. 1, pp. 61-73.
- Loschky, L.C., Larson, A.M., Magliano, J.P. and Smith, T.J. (2015), "What would jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension", *Plos One*, Vol. 10 No. 11, p. e0142474.
- Mattke, J., Maier, C., Reis, L. and Weitzel, T. (2021), "In-app advertising: a two-step qualitative comparative analysis to explain clicking behavior", *European Journal of Marketing*, Vol. 55 No. 8, pp. 2146-2173.
- Miller, G.A. (1956), "The magical number seven, plus or minus two: some limits on our capacity for processing information", *Psychological Review*, Vol. 63 No. 2, pp. 81-97.
- Morton, J. (1967), "A singular lack of incidental learning", *Nature*, Vol. 215 No. 5097, pp. 203-204.
- Myers, S.D., Deitz, G.D., Huhmann, B.A., Jha, S. and Tataru, J.H. (2020), "An eye-tracking study of attention to brand-identifying content and recall of taboo advertising", *Journal of Business Research*, Vol. 111, pp. 176-186.
- Nummenmaa, L., Hyönä, J. and Calvo, M.G. (2006), "Eye movement assessment of selective attentional capture by emotional pictures", *Emotion*, Vol. 6 No. 2, pp. 257-268.
- Orquin, J.L. and Holmqvist, K. (2018), "Threats to the validity of eye-movement research in psychology", *Behavior Research Methods*, Vol. 50 No. 4, pp. 1645-1656.
- Orquin, J.L. and Loose, S.M. (2013), "Attention and choice: a review on eye movements in decision making", *Acta Psychologica*, Vol. 144 No. 1, pp. 190-206.
- Palermo, R. and Rhodes, G. (2007), "Are you always on my mind? A review of how face perception and attention interact", *Neuropsychologia*, Vol. 45 No. 1, pp. 75-92.
- Payne, S.J., Howes, A. and Reader, W.R. (2001), "Adaptively distributing cognition: a decision-making perspective on human-computer interaction", *Behaviour and Information Technology*, Vol. 20 No. 5, pp. 339-346.
- Pieters, R. and Wedel, M. (2004), "Attention capture and transfer in advertising: brand, pictorial, and text-size effects", *Journal of Marketing*, Vol. 68 No. 2, pp. 36-50.
- Pieters, R. and Wedel, M. (2007), "Goal control of attention to advertising: the Yarbus implication", *Journal of Consumer Research*, Vol. 34 No. 2, pp. 224-233.
- Pieters, R., Wedel, M. and Batra, R. (2010), "The stopping power of advertising: measures and effects of visual complexity", *Journal of Marketing*, Vol. 74 No. 5, pp. 48-60.
- Posner, M.I. and Cohen, Y. (1984), "Components of visual orienting", in Bouma, H. and Bouwhuis, D.G. (Eds), *Attention and Performance X: Control of Language Processes*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 531-556.
- Pourtois, G., Schettino, A. and Vuilleumier, P. (2013), "Brain mechanisms for emotional influences on perception and attention: what is magic and what is not", *Biological Psychology*, Vol. 92 No. 3, pp. 492-512.

-
- Ravaja, N., Somervuori, O. and Salminen, M. (2013), "Predicting purchase decision: the role of hemispheric asymmetry over the frontal cortex", *Journal of Neuroscience, Psychology, and Economics*, Vol. 6 No. 1, pp. 1-13.
- Rayner, K. (1998), "Eye movements in reading and information processing: 20 years of research", *Psychological Bulletin*, Vol. 124 No. 3, pp. 372-422.
- Rayner, K., Smith, T.J., Malcolm, G.L. and Henderson, J.M. (2009), "Eye movements and visual encoding during scene perception", *Psychological Science*, Vol. 20 No. 1, pp. 6-10.
- Reeves, B. and Nass, C.I. (1996), *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*, Cambridge University Press, New York, NY.
- Reeves, B., Thorson, E., Rothschild, M.L., McDonald, D., Hirsch, J. and Goldstein, R. (1985), "Attention to television: intrastimulus effects of movement and scene changes on alpha variation over time", *International Journal of Neuroscience*, Vol. 27 Nos 3/4, pp. 241-255.
- Rogers, Y. (2004), "New theoretical approaches for human-computer interaction", *Annual Review of Information Science and Technology*, Vol. 38 No. 1, pp. 87-143.
- Romaniuk, J. (2009), "The efficacy of brand-execution tactics in TV advertising, brand placements and internet advertising", *Journal of Advertising Research*, Vol. 49 No. 2, pp. 143-150.
- Rosbergen, E., Pieters, R. and Wedel, M. (1997), "Visual attention to advertising: a segment-level analysis", *Journal of Consumer Research*, Vol. 24 No. 3, pp. 305-314.
- Rosenholtz, R. (1999), "A simple saliency model predicts a number of motion popout phenomena", *Vision Research*, Vol. 39 No. 19, pp. 3157-3163.
- Rosenholtz, R., Li, Y. and Nakano, L. (2007), "Measuring visual clutter", *Journal of Vision*, Vol. 7 No. 2, pp. 3157-3163.
- Rösler, L., Rubo, M. and Gamer, M. (2019), "Artificial faces predict gaze allocation in complex dynamic scenes", *Frontiers in Psychology*, Vol. 10 No. 2877, pp. 1-10.
- Ross, N.M. and Kowler, E. (2013), "Eye movements while viewing narrated, captioned, and silent videos", *Journal of Vision*, Vol. 13 No. 4, pp. 1-19.
- Rubo, M. and Gamer, M. (2018), "Social content and emotional valence modulate gaze fixations in dynamic scenes", *Scientific Reports*, Vol. 8 No. 1, pp. 1-11.
- Rumpf, C., Boronczyk, F. and Breuer, C. (2020), "Predicting consumer gaze hits: a simulation model of visual attention to dynamic marketing stimuli", *Journal of Business Research*, Vol. 111, pp. 208-217.
- Russell, C.A. (2002), "Investigating the effectiveness of product placements in television shows: the role of modality and plot connection congruence on brand memory and attitude", *Journal of Consumer Research*, Vol. 29 No. 3, pp. 306-318.
- Simmonds, L., Bogomolova, S., Kennedy, R., Nenycz-Thiel, M. and Bellman, S. (2020), "A dual-process model of how incorporating audio-visual sensory cues in video advertising promotes active attention", *Psychology and Marketing*, Vol. 37 No. 8, pp. 1057-1067.
- Simons, R.F., Detenber, B.H., Roedema, T.M. and Reiss, J.E. (1999), "Emotion processing in three systems: the medium and the message", *Psychophysiology*, Vol. 36 No. 5, pp. 619-627.
- Simons, R.F., Detenber, B.H., Cuthbert, B.N., Schwartz, D.D. and Reiss, J.E. (2003), "Attention to television: alpha power and its relationship to image motion and emotional content", *Media Psychology*, Vol. 5 No. 3, pp. 283-301.
- Smith, K.C. and Abrams, R.A. (2018), "Motion onset really does capture attention", *Attention, Perception, and Psychophysics*, Vol. 80 No. 7, pp. 1775-1784.
- Song, J., Newton, O.B., Fiore, S.M., Pittman, C. and LaViola, J.J., Jr (2019), "Examining training comprehension and external cognition in evaluations of uncertainty visualizations to support decision making", *Proceedings of the Human Factors and Ergonomics Society 2019 Annual Meeting*, SAGE Publications, Los Angeles, CA, pp. 1654-1658.

- Stewart, D.W. and Furse, D.H. (1986), *Effective Television Advertising: A Study of 1000 Commercials*, Lexington Books, Lexington, MA.
- Tatler, B.W., Hayhoe, M.M., Land, M.F. and Ballard, D.H. (2011), "Eye guidance in natural vision: reinterpreting salience", *Journal of Vision*, Vol. 11 No. 5, pp. 1-23.
- Teixeira, T., Wedel, M. and Pieters, R. (2010), "Moment-to-moment optimal branding in TV commercials: preventing avoidance by pulsing", *Marketing Science*, Vol. 29 No. 5, pp. 783-804.
- Theeuwes, J. (1991), "Cross-dimensional perceptual selectivity", *Attention, Perception, and Psychophysics*, Vol. 50 No. 2, pp. 184-193.
- Treisman, A.M. and Gelade, G. (1980), "A feature-integration theory of attention", *Cognitive Psychology*, Vol. 12 No. 1, pp. 97-136.
- Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H. and Livingstone, M.S. (2006), "A cortical region consisting entirely of face-selective cells", *Science*, Vol. 311 No. 5761, pp. 670-674.
- Valiyamattam, G.J., Katti, H., Chaganti, V.K., O'Haire, M.E. and Sachdeva, V. (2020), "Do animals engage greater social attention in autism? An eye tracking analysis", *Frontiers in Psychology*, Vol. 11 No. 727, pp. 1-10.
- Van der Burg, E., Cass, J. and Theeuwes, J. (2019), "Changes (but not differences) in motion direction fail to capture attention", *Vision Research*, Vol. 165, pp. 54-63.
- Van Der Lans, R., Pieters, R. and Wedel, M. (2008), "Competitive brand salience", *Marketing Science*, Vol. 27 No. 5, pp. 922-931.
- Von Wartburg, R., Ouerhani, N., Pflugshaupt, T., Nyffeler, T., Wurtz, P., Hügli, H. and Müri, R.M. (2005), "The influence of colour on oculomotor behaviour during image perception", *NeuroReport*, Vol. 16 No. 14, pp. 1557-1560.
- Wang, Z., Ren, J., Zhang, D., Sun, M. and Jiang, J. (2018), "A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos", *Neurocomputing*, Vol. 287, pp. 68-83.
- Wedel, M. and Pieters, R. (2006), "Eye tracking for visual marketing", *Foundations and Trends® in Marketing*, Vol. 1 No. 4, pp. 231-320.
- Wedel, M. and Pieters, R. (2017), "A review of eye-tracking research in marketing", *Review of Marketing Research*, pp. 123-147.
- Wedel, M., Pieters, R. and van der Lans, R. (2019), "Eye tracking methodology for research in consumer psychology", in Kardes, F.R., Herr, P.M. and Norbert Schwartz, N. (Eds), *Handbook of Research Methods in Consumer Psychology*, Routledge, New York, NY, pp. 276-292.
- Williams, C.C. and Castelano, M.S. (2019), "The changing landscape: high-level influences on eye movement guidance in scenes", *Vision*, Vol. 3 No. 3, pp. 33-52.
- Wolfe, J.M. and Horowitz, T.S. (2004), "What attributes guide the deployment of visual attention and how do they do it?", *Nature Reviews Neuroscience*, Vol. 5 No. 6, pp. 495-501.
- Yang, Y., Li, B., Li, P. and Liu, Q. (2018), "A two-stage clustering based 3d visual saliency model for dynamic scenarios", *IEEE Transactions on Multimedia*, Vol. 21 No. 4, pp. 809-820.
- Yarbus, A.L. (1967), *Eye Movements and Vision*, Plenum press, New York, NY.
- Yegiyan, N.S. and Lang, A. (2010), "Processing central and peripheral detail: how content arousal and emotional tone influence encoding", *Media Psychology*, Vol. 13 No. 1, pp. 77-99.
- Yegiyan, N.S. and Yonelinas, A.P. (2011), "Encoding details: positive emotion leads to memory broadening", *Cognition and Emotion*, Vol. 25 No. 7, pp. 1255-1262.

Appendix 1. Additional methodology details

Participants

The total sample consisted of 49 members of a Southwestern US audience panel, all with normal or corrected-to-normal vision. Data from 12 participants with poor eye-tracking calibration were deleted. The final sample of 37 participants is comparable in size with other studies using a within-participants, repeated measures design (Guerreiro *et al.*, 2015: $N = 41$; Ravaja *et al.*, 2013: $N = 33$) and had 80% power of detecting a significant ($p < 0.05$) small effect-size difference ($d = 0.2$, $r = 0.1$, standardized $\beta = 0.05$) between any of the 34 advertisements seen by each participant (Faul *et al.*, 2009). Increasing the power of the data, the number of repeated measures was even higher than 34 per participant; each advertisement contained several scenes and each scene depicted one or more AOIs.

Apparatus

Tobii T60 eye-trackers had a sampling rate of 60 Hz and a tracking accuracy of 0.5 degrees. Each unit consisted of a 24-inch video monitor with infra-red eye-tracking sensors located under the screen between the monitor's speakers. The software used to collect and analyse participants' eye movement data was Attention Tool 5.1 by iMotions Global®.

Video advertisement selection

Unfamiliar, out-of-market (Australian) video advertisements were presented to the sample of US participants so that eye movements could be attributed to the video content rather than differences in individuals' prior exposure (Rosbergen *et al.*, 1997). Australia and the USA are close culturally (House *et al.*, 2010) and the creative content was in English but was novel, even for known brands like Ford and Kellogg's. The 30-second advertisements were recorded at randomly selected times from the three biggest Australian free-to-air television networks across one week of peak (winter) viewing.

The final sample of 34 advertisements (for 34 different brands) spanned a mix of product and service categories, including automobiles, finance, food, home and personal care products and retailers. The programme included five advertising breaks, each containing seven advertisements (i.e. a total of 35 ads, including one advertisement that was not analysed because of its different aspect ratio).

Coding areas of interest

A key step in identifying AOIs in dynamic video is to break the video into "scenes", so that during each scene the number of potential AOIs stayed constant. The entry of a new object on the screen potentially introduces a new AOI, necessitating the start of a new scene (Abrams and Christ, 2003; Yantis and Hillstrom, 1994). A new scene might begin with each new frame; in a 25 frames-per-second video, this would occur every 0.04 of a second. We used an empirical cut-off for minimum scene duration, based on our data, of 0.36 s (i.e. nine video frames) to exclude fixations from the previous scene (Le Meur *et al.*, 2007), as the longest saccade prompted by the previous scene would have lasted less than 0.2 of a second (Fischer and Ramsperger, 1984). Representative snapshots were taken of each scene, for the 11 to 28 scenes in each advertisement (where most scenes were several seconds long), from which the AOIs were identified and coded. Throughout the duration of a scene, the IBBs around each AOI in the scene could move and change shape to keep track of the objects they defined. The software counted fixations and fixation durations for the entirety of the advertisements and not just for the key frames that were used to identify the AOIs present in each scene.

Identifying areas of interest

Several criteria were used to identify important AOIs in the scenes. AOI importance was suggested by salience dimensions such as size and brightness and by dynamic-attention dimensions such as imminence (centrality). When multiple AOIs were present, priority was given to AOIs present across multiple scenes, as these would have greater task relevance for following the advertising story (Lang and Bailey, 2015). AOIs were also considered task relevant if they depicted the advertised product or included graphics (e.g. charts), text (e.g. slogans) or the brand's logo. AOIs with potential motivational relevance included animals, faces and eyes. A new dimension considered for identifying AOIs was stability (fleeting vs stable). Movement in any direction (Itti, 2004), or growing or shrinking in size (Franconeri and Simons, 2003), was coded as fleeting = 1 (as opposed to stable = 0). Certain actions that might have been coded as task relevant were coded as fleeting because they typically involved movement, such as interacting with the product, or when AOIs were speaking (e.g. a face or person). Table A1 lists definitions of the codes used in the analysis, with their corresponding means (% prevalence for binary codes) and standard deviations.

Figure A1 shows a coded snapshot from one of the advertisements. In the scene depicted, the seven AOIs are (1) the product, (2) text (slogan), (3) character's hands (interacting with product), (4) bowl of food, (6) fork (all outlined by IBBs) and finally (7) the background. Visual clutter was coded as 1 (clutter present) if the scene included four or more AOIs (like the scene in Figure A1 does), otherwise 0.

Two people independently coded five of the 34 advertisements to check the reliability of the coding scheme. Krippendorff's alpha (De Swert, 2012) confirmed that all codes were reliable, exceeding 0.67 (range 0.70 to 1.00). Codes that did not appear in at least four of the 34 commercials were combined with another code if appropriate. For this reason, the colours green and purple were combined with their more saturated counterparts of dark green and dark purple.

Data collection procedure

Participants viewed the content in an individual viewing lab. They sat at a desk approximately 60 to 70 cm away from the T60 eye-tracking monitor. Once the eye-gaze equipment was calibrated to the participant and instructions were provided, the programme began. After the participant had finished watching the programme, they were compensated, debriefed and dismissed.

Analysis

Using the IBB drawing tool in Attention Tool 5.1, the identified AOIs were coded and then fixation durations in milliseconds were automatically calculated for each AOI, across the duration of the advertisement, with a 100 ms fixation threshold (Duchowski, 2007; Rayner, 1998). The durations of all participants' fixations inside an AOI were summed to measure fixation duration.

Of the 99,715 potential observations in the data set, 9,209 (9%) had missing data because of the loss of eye-tracking, leaving a total of 90,506 fixation observations. Of these, 51,358 (57%) had durations equal to 100 ms or more, coded as fixation = 1; the remaining 39,148 observations (39%) were coded as fixation = 0. Observations with fixation = 0 were deleted from the duration variable ($N = 51,358$, median = 584ms, range 100 ms to 6,997ms). The duration variable was natural-log transformed ($M = 6.32$ and $SD = 0.87$) to normalise its positively skewed distribution (Barr, 2008). There was no need to delete outliers, as the log-transformed minimum and maximum observations were less than three standard deviations from the mean.

Fixed effects (participant constants) were used to accommodate the correlations between data from the same participant (Barr, 2008). Logistic regression was used for the binary fixation variable and general linear model analysis for the normally distributed log-transformed duration variable.

Independent variable	Definition	<i>M</i> ^a	<i>SD</i>	Influence of dynamic content
<i>Static saliency variables</i>				
Size	Area of interest (AOI) size (area) in pixels as a percentage of total screen area (0–1)	0.22	0.30	161
Brightness	Brightness histogram median [0-255 (transformed to 0–1 for the regression analyses)]	103.8	56.53	
Clutter	4 or more (maximum = 6) foreground AOIs = 1, else 0	0.62	0.49	
<i>Colour variables</i> ^b				
Red	Red (R 255, G 0, B 0) as a % of AOI area (0–1, centred)	0.003	0.03	
Dark red	R128, G 0, B 0 (0–1, centred)	0.02	0.07	
Orange	R 255, G 165, B 0 (0–1, centred)	0.02	0.1	
Dark orange	R 128, G 83, B 0 (0–1, centred)	0.16	0.16	
Yellow	R 255, G 255, B 0 (0–1, centred)	0.002	0.02	
Dark yellow	R 128, G 128, B 0 (0–1, centred)	0.01	0.04	
Green (including dark green)	R 0, G 128 or 255, B 0 (0–1, centred)	0.003	0.03	
Blue	R 0, G 0, B 255 (0–1, centred)	0.002	0.02	
Dark blue	R 0, G 0, B 128 (0–1, centred)	0.03	0.1	
Purple (including dark purple)	R 128 or 255, G 0, B 128 or 255 (0–1, centred)	0.005	0.03	
Grey	R 128, G 128, B 128 (0–1, centred)	0.39	0.26	
White	R 255, G 255, B 255 (0–1, centred)	0.13	0.21	
<i>Location</i>				
Central location	Fixation in middle ninth of screen = 1, else 0	0.39	0.49	
<i>Motivational relevance</i> ^c				
Animal	AOI is animal = 1, else 0	0.01	0.08	
Face	AOI is face = 1, else 0	0.12	0.33	
Eyes	AOI is eye(s) = 1, else 0	0.04	0.20	
<i>Task relevance</i>				
Graphic	AOI is graphic = 1, else 0	0.04	0.20	
Text	AOI is text = 1, else 0	0.12	0.33	
Product	AOI is product = 1, else 0	0.15	0.36	
Visual branding	AOI is visual branding = 1, else 0	0.01	0.09	
Logo	AOI is logo = 1, else 0	0.02	0.14	
<i>Stability</i>				
Movement	Area of interest [AOI] is moving/changing size = 1, else 0	0.70	0.46	
Time	Time on-screen in seconds [short = 1 (less than mean), long = 0]	1.71	0.97	
Interacting with product	AOI is interacting with product = 1, else 0	0.04	0.19	
Speaking	AOI speaking = 1, else 0	0.05	0.21	

Notes: (a) Mean (*M*) represents the percentage of scenes with the variable present for binary (present/absent) variables; (b) screenshots were reduced to 15 colours [six primary, six secondary (darker) and three neutral] and the area occupied by each colour was divided by total AOI area, represented as a percentage; (c) for motivational relevance, task relevance and stability, the presence of any of the variables associated with the dimension indicated that the dimension was present

Table A1. Independent variable codes, means and standard deviations

Figure A1.
Areas of interest
outlined with
intelligent bounding
boxes



Hierarchical regression was used to measure the amount of additional variance explained by each successive block of variables. The first block consisted of the individual-level fixed effects and explained less than 1% of variation in the fixation model and 6.9% in the fixation duration model (Table A3). These models were not affected by multicollinearity, as the largest VIF was 4.07 for the three-way interaction of Imminence \times Motivational Relevance \times Stability, which is less than the traditional limit of 10 (Petter *et al.*, 2007).

Two additional models were estimated (results available on request from the authors) to confirm that the reported results in Table A3 were not affected by potential confounding effects associated with the individual advertisements (e.g. because of brand or product effects). In one of these additional models, for fixation duration, these potential confounding effects were controlled for by adding more fixed effects (i.e. one for each advertisement). The results of this new model were substantially the same as the reported results in Table A3. Adding these fixed effects explained more variance in Block 1 (9.6%), compared to the results in Table A3 (6.9%), but reduced the variance explained by the Stability factor in Block 7 (1.6% vs 2.5%). All the significant results in the fixation duration model reported in Table A2 were significant and in the same direction in the new, additional fixed effects model, except for the results for one colour, dark yellow, which now had an insignificant negative effect ($b = -0.12$, $SE = 0.09$, $\beta = -0.01$, $t = -1.31$ and $p = 0.19$).

We also tried adding advertisement fixed effects to the logistic regression model of fixation, but this model could not be estimated. Instead, we used a generalized mixed regression approach with random effects for the interaction between participant and ad. Preliminary analyses showed that this specification explained more variance than a model with separate random effects for each participant and each ad. Again, all the significant results for the fixation model in Table A3 were significant and, in the same direction, in the new model. These tests suggest that the results in Table A3, which use the same fixed-effects model for both fixation and fixation duration, were not affected by any potential confounding effects of the brands and products used in the 34 advertisements.

Variable	Fixation	Fixation duration
<i>Saliency variables</i>		
Size (0–1, centred)	0.45***	0.37***
Brightness median (0–1, centred)	–0.05***	–0.04***
Clutter	–0.17***	–0.03***
<i>Colour variables</i>		
Red percent (0–1, centred)	0.01	–0.02***
Dark red percent (0–1, centred)	–0.01	–0.03***
Orange percent (0–1, centred)	–0.02***	0.03***
Dark orange percent (0–1, centred)	–0.01***	–0.04***
Yellow percent (0–1, centred)	0.03***	0.03***
Dark yellow percent (0–1, centred)	0.03***	0.04***
Green percent (0–1, centred)	0.03***	0.05***
Blue percent (0–1, centred)	0.02***	0.004
Dark blue percent (0–1, centred)	–0.002	–0.01
Purple percent (0–1, centred)	–0.02***	–0.01*
Grey percent (0–1, centred)	–0.05***	–0.07***
White percent (0–1, centred) ^a	–0.02***	0.004
<i>Imminence</i>		
Central location	0.01	–0.20***
<i>Motivational relevance</i>		
Animal	0.03***	0.001
Face	0.001	–0.10***
Eyes	–0.05***	–0.02***
<i>Task relevance</i>		
Graphic	–0.08***	–0.06***
Text	–0.08***	–0.07***
Product	–0.02***	–0.10***
Visual branding	0.02***	0.003
Logo	–0.03***	–0.04***
<i>Stability</i>		
Movement	–0.11***	–0.13***
Time on screen (seconds)	0.10***	0.31**
Interacting with product	–0.09***	–0.06***
Speaking	0.02***	–0.01

Notes: Fixation ($N = 90,506$) was dichotomous (1/0) and Duration (sum of fixations ≥ 100 ms, $N = 51,358$) was normalised by a natural log transformation; ^a This variable was used as the reference category in the regression analyses; * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$

Table A2.
Zero-order
correlations of
constituent variables
with dependent
variables

Table A3.
Hierarchical
regression models of
fixation and duration

Independent variable	Fixation		Dependent variable		Fixation duration	
	<i>b</i>	<i>OR</i>	<i>Sig</i>	<i>b</i>	<i>b</i>	<i>Sig</i>
<i>Block 1: Fixed effects</i>						
Intercept	1.13	$R^2 = 0.7\%$, $\Delta R^2 = 0.7\%$	***	6.75	$R^2 = 6.9\%$, $\Delta R^2 = 6.9\%$	***
<i>Block 2: Solence variables</i>						
Size (0-1, centred)	5.40	$R^2 = 32.2\%$, $\Delta R^2 = 31.5\%$	***	0.62	$R^2 = 21.2\%$, $\Delta R^2 = 14.3\%$	***
Brightness median (0-1, centred)	-0.01	221.21	***	-0.12	-0.03	***
Clutter	-0.36	0.70	***	0.01	0.01	***
<i>Block 3: Colour variables</i>						
Red percent (0-1, centred)	0.76	$R^2 = 32.6\%$, $\Delta R^2 = 0.4\%$	**	-0.35	$R^2 = 21.5\%$, $\Delta R^2 = 0.4\%$	**
Dark red percent (0-1, centred)	0.39	2.14	***	0.06	-0.01	***
Orange percent (0-1, centred)	-0.61	1.47	***	0.02	0.005	***
Dark orange percent (0-1, centred)	0.51	0.54	***	0.02	0.002	***
Yellow percent (0-1, centred)	5.01	149.65	***	1.01	0.04	***
Dark yellow percent (0-1, centred)	0.03	1.04	***	0.27	0.01	**
Green percent (0-1, centred)	0.39	1.47	***	0.21	0.01	*
Blue percent (0-1, centred)	-0.08	0.92	***	0.03	0.001	*
Dark blue percent (0-1, centred)	-0.22	0.80	**	-0.24	-0.03	***
Purple percent (0-1, centred)	0.46	1.58	***	0.50	0.02	***
Grey percent (0-1, centred)	0.17	1.18	***	-0.01	-0.002	***
White percent (0-1, centred) ^a	-	-	-	-	-	-
<i>Block 4: Imminence</i>						
Central location (1/0)	0.41	$R^2 = 34.6\%$, $\Delta R^2 = 2.0\%$	***	-0.20	$R^2 = 21.5\%$, $\Delta R^2 = 0.01\%$	*
<i>Block 5: Motivational relevance</i>						
Motivational relevance present (1/0)	-0.21	$R^2 = 34.6\%$, $\Delta R^2 = 0.1\%$	***	-0.24	$R^2 = 21.6\%$, $\Delta R^2 = 0.1\%$	***
<i>Block 6: Task relevance</i>						
Task Relevance present (1/0)	-0.25	$R^2 = 34.7\%$, $\Delta R^2 = 0.1\%$	***	-0.36	$R^2 = 21.7\%$, $\Delta R^2 = 0.1\%$	***
<i>Block 7: Stability</i>						
Fleeting (= 1, Stable = 0)	-0.49	$R^2 = 34.8\%$, $\Delta R^2 = 0.1\%$	***	-0.48	$R^2 = 24.2\%$, $\Delta R^2 = 2.5\%$	***
<i>Block 8: Interaction effects</i>						
Imminence × Motivational Relevance × Fleeting	0.28	$R^2 = 35.1\%$, $\Delta R^2 = 0.3\%$	***	0.03	$R^2 = 25.2\%$, $\Delta R^2 = 1.0\%$	***
Imminence × Task Relevance × Fleeting	0.66	1.33	***	0.47	0.13	***
		1.94	***		0.17	***

Notes: Fixation ($N = 90,506$) was dichotomous (1/0) and Duration (sum of fixations ≥ 100 ms, $N = 51,358$) was normalised by a natural log transformation; *b* = regression coefficient, *OR* = odds ratio, *Sig* = significance and β = standardised beta; Nagelkerke R^2 reported for the logistic regression results (fixation). ^a This variable was used as the reference category; * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$

Appendix 2. Additional detailed results

Saliency theory and dimensions

Every block, including the saliency variables block, added significant R^2 to the fixation duration model, according to F -tests. The positive effect of size, and the negative effect of brightness, reflected the signs of their zero-order correlations in [Table A2](#). The insignificant effect of clutter on fixation duration in [Table A3](#) contrasted with its significant negative correlation with fixation duration ([Table A2](#)). This may have been because the regression model controlled for the effects of other variables.

The colour variables were added as a separate block to test the explanatory value of the authors' novel colour coding method, extending the colour palette beyond prior research. The positive effect of yellow on attracting fixation might potentially be explained by its rarity in these data (associated with just 0.2% of AOIs) but similarly rare colours had significant negative effects [red (0.3% of AOIs) on fixation duration] or no significant effects [blue (0.2% of AOIs) on fixation and fixation duration]. Orange (2% of AOIs) and dark blue (3% of AOIs) had significant negative effects on fixation and/or fixation duration. All other effects were positive, although not always significant for both fixation and fixation duration (e.g. dark red).

These colour results also largely reflected the raw correlations in [Table A2](#), which included several negative correlations. Grey, the most common colour (present in 39% of AOIs), had negative correlations with fixation and fixation duration, but a significant positive effect on fixation after controlling for other variables in the model ([Table A3](#)). Grey's negative effect on fixation duration was not significant in the regression model.

Dynamic attention theory and dimensions

Imminence. Only 40% of AOIs were located centrally ([Table A1](#)), despite centrality being a criterion for identifying important AOIs. The regression results replicated the directions of imminence's correlations with fixation and fixation duration ([Table A2](#)).

Motivational relevance. Motivational relevance was mainly defined by the presence of a face, which was the case for 12% of AOIs ([Table A1](#)). The negative regression results reflected the correlations in [Table A2](#). Only animals, which were associated with just 1% of AOIs, had a significant positive correlation with attention (specifically, fixation). All other correlations were either negative or not significant. The negative correlation for eyes accords with prior research, showing that while eyes attract attention in static pictures, in dynamic video, attention is just as likely to be attracted by other more relevant areas of the face, such as the mouth when the person is talking (Vö *et al.*, 2012).

Task relevance. Task relevance was mainly defined by the presence of a product, which was the case for 15% of AOIs ([Table A1](#)). The negative regression results reflected the negative task-relevance correlations in [Table A2](#), except for visual branding, which correlated positively with fixation.

Stability. The fleeting versus stable dimension was mainly defined by movement being present ([Table A1](#)), which was associated with 70% of AOIs. Short screen-time was also a defining variable, as, using a median split, half of our AOIs were defined as fleeting because their time on-screen was less than the mean/median (1.71 s). A small number of fleeting AOIs were defined by the presence of interaction with the product (4%) or speaking (5%). The negative regression results for fleeting versus stable stimuli reflected the small but significant negative correlations between movement and fixation and fixation duration ([Table A2](#)). Fixation duration was positively correlated with AOI time on-screen (averaged across the scenes it appeared in) ($r = 0.31$), so fixation duration was negatively

correlated with fleeting AOIs with short time on-screen. Similarly, fixation was positively correlated with time on-screen, so negatively correlated with short time on-screen (fleeting stimuli). Interacting with the product also had negative correlations with fixation and fixation duration.

References

- Barr, D.J. (2008), "Analyzing 'visual world' eye tracking data using multilevel logistic regression", *Journal of Memory and Language*, Vol. 59 No. 4, pp. 457-474.
- De Swert, K. (2012), "Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha", Working Paper, University of Amsterdam, 1 February.
- Duchowski, A. (2007), *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag New York, NY.
- Faul, F., Erdfelder, E., Buchner, A. and Lang, A.-G. (2009), "Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses", *Behavior Research Methods*, Vol. 41 No. 4, pp. 1149-1160.
- Fischer, B. and Ramsperger, E. (1984), "Human express saccades: extremely short reaction times of goal directed eye movements", *Experimental Brain Research*, Vol. 57, pp.191-195.
- Franconeri, S.L. and Simons, D.J. (2003), "Moving and looming stimuli capture attention", *Attention, Perception, and Psychophysics*, Vol. 65 No. 7, pp. 999-1010.
- House, R.J., Quigley, N.R. and de Luque, M.S. (2010), "Insights from project globe: extending global advertising research through a contemporary framework", *International Journal of Advertising*, Vol. 29 No. 1, pp. 111-139.
- Le Meur, O., Le Callet, P. and Barba, D. (2007), "Predicting visual fixations on video based on low-level visual features", *Vision Research*, Vol. 47 No. 19, pp. 2483-2498.
- Petter, S., Straub, D. and Rai, A. (2007), "Specifying formative constructs in information systems research", *MIS Quarterly*, Vol. 31 No. 4, pp. 623-656.
- Võ, M.L.H., Smith, T.J., Mital, P.K. and Henderson, J.M. (2012), "Do the eyes really have it? Dynamic allocation of attention when viewing moving faces", *Journal of Vision*, Vol. 12 No. 13, pp. 1-14.
- Yantis, S. and Hillstrom, A.P. (1994), "Stimulus-driven attentional capture: evidence from equiluminant visual objects", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 20 No. 1, p. 95.

Corresponding author

Steven Bellman can be contacted at: Steven.Bellman@marketingscience.info