Consumer Perception at Point of Purchase: Evaluating Proposed Package Designs in an Eye-tracking Lab

Erik Wästlund, Poja Shams, Martin Löfgren, Lars Witell & Anders Gustafsson
CTF-Service research center, Karlstad University.

Keywords
Eye-tracking, point-of-purchase marketing, FMCG, Package design.

Abstract
In today’s retail environment, consumer products are increasingly competing for customers’ attention. Research has shown that 60–80% of purchasing decisions are influenced in-store. Thus, packaging that stands out from competitors gains a competitive advantage. This study investigates the use of eye-tracking as a method to evaluate and design packaging with better Point-of-Purchase qualities. An eye-tracking laboratory was used and shoppers were recruited for three rounds of experiments. In total, 128 participants were recruited in order to assess the potential of eye-tracking. Results show that, when taking some methodological constraints into account, eye-tracking complements traditional methodologies with further insights when investigating the Point-of-Purchase qualities of packaging.

Introduction
In the retail environment, fast moving consumer goods (FMCGs) are increasingly competing for customers’ attention. In this context, packaging has been acknowledged as a silent salesperson (Pilditch, 1973), which not only involves capturing shoppers’ attention, but maintaining it amid the visual clamour of competing products and brands (Judd, Aalders, & Melis, 1989). Finding a product on the shelf has become a challenge, not just because of the large number of options, but also because of shared product attributes and copy-cat products that are available at the Point of Purchase (POP).

Underwood and Klein (2002) have shown that packaging becomes increasingly important for decisions made at the point of purchase, relative to other communication tools, because of its easy availability. Additionally, 60–80% of all brand purchasing decisions are made in-store (POPAI, 1996). Customers take only seconds to choose between products while shopping (Judd et al., 1989) and 90% of consumer decisions are made simply by looking at the front of the package without even touching the actual product (Urbany, Dickson, & Kalapurakal, 1996). Since the package often projects the customer’s first impression of a brand’s quality or value (Harckham, 1989), the ability of a package to stand out from its competitors becomes a competitive advantage. This raises the question of how the decision-making processes of customers can be understood in order to provide packaging with better POP qualities.

Although there is a limited amount of research on process of customer decision in a store we know that 83% of all the information that the brain processes is received through the eyes (Nordfält, 2007). Hence if we can capture the decision process by understanding what a consumer is looking at we build more knowledge about what the selling features are for a package. The aim of this paper is to investigate and describe how eye tracking methodology can be used to evaluate packages for the purpose of making sure that the packages are seen by consumers. In this process we also build a better understanding of how consumers perceive shelves during their shopping. Three rounds of experiments, involving 128 participants, were conducted for the present study in order to assess eye-tracking as a potential factor to be incorporated in the design of packages. The results show that the eye-tracking methodology offers insights in addition to traditional methodologies when investigating the POP qualities of packaging.
The Role of Packaging at the Point of Purchase

Packaging is a critical factor in the consumer decision-making process because it actually communicates to the consumer in the store (Silayoi & Speece, 2007). The package stimulates purchasing behavior because it is a medium of attention, information, and aesthetics (Bech-Larsen, 1996). Packaging that captures the customer’s attention facilitates quick in-store decision-making (Silayoi & Speece, 2004). The results of previous research into packaging and package design suggest that packaging is an influential element at the POP, and that it has a pervasive impact on customers during their purchasing decision process (Orth & Malkewitz, 2008).

Attracting customers’ attention, however, can be difficult because of the large number of competing stimuli (products) in a store, and the fact that most of these stimuli are ignored by customers who shop habitually (Underwood, Klein, & Burke, 2001). Nonetheless, once the shopper’s attention has been caught, the features of the packaging can serve to underline the uniqueness and originality of the product (Silayoi & Speece, 2007). Furthermore, visual packaging elements play a major role, representing the product to many consumers, especially during low-involvement shopping or rushed shopping situations (Silayoi & Speece, 2004).

As the customer’s eye wanders across the products on the shelf, some packages stand out and attract the customer’s attention. As it is very difficult to decouple attention from gaze direction during complex cognition (Rayner, 1998) eye location can be used as an indicator of overt attention. Although it is difficult to accurately predict purchasing behavior based solely on overt attention, it is possible to conclude at the most basic level that unseen is unsold. The question then is, ‘How can packages be designed and evaluated in order to make sure they are seen?’

The Eye-tracking Methodology

Eye-tracking is a method for measuring eye movements while viewing a digital or physical stimulus. There are several ways to measure eye movement, the most common of which is by using video-based combined pupil/corneal reflection (Duchowski, 2007). The human eye moves constantly until the focus is set on an object. The movements are called saccades and the focus point is called a fixation. Saccades last between 10 and 100 ms and during this time the eyes do not send information about the stimulus in question to the brain (Duchowski, 2007). After the saccades, the eyes will rest on an object (fixation) and the brain will start processing the stimulus. Switching between saccades and fixations forms a search pattern (Goldberg & Kotval, 1999).

The data from an eye-tracking study can either be classified as process-based or performance-based. Performance-based data are generated via search tasks given to the participant, such as looking for a specific product (Oreo), product category (cookie), product attribute (blue and cylindrical in shape), or product property (everyday brand of cookie). Measurements such as response time and number of clicks can be quantified by the active involvement of the participant. Process-based results are based on the eye movements that occur while the respondent is performing the task. This provides information about attention on the packages that are being looked at, which can in turn; provide the information needed to solve the task. Analyzing gaze data makes it possible to surmise which parts of a package facilitate or hamper the decision-making process.

Visual impact or saliency, the subjective perceptual quality of an object, determines the effectiveness of detecting it during a visual search (Yantis & Egeth, 1999). Typical measurements of visual impact include the time to first fixation and the percentage noting, that is, the percentage of respondents who actually look at a product. If a given brand or package is not salient, it will not be considered as an alternative at the point of sale (Holman & Hecker, 1983). The visual impact of a package is an important component of its findability,
or how easy it is to find when looked for, as opposed to merely standing out when a customer is looking at a shelf in general. There are two uncertainties involved in the process of finding a target product: location uncertainty and identity uncertainty (Van der Lans, Pieters, & Wedel, 2008). Location uncertainty is the uncertainty of finding an object on the shelf, while identity uncertainty is the uncertainty of an identified object being a target or a distracter. Response time and accuracy are two typical measurements of findability.

Visual impact and findability can be affected by both top-down and bottom-up information. Top-down information is memories and knowledge gained from previous experiences. Bottom-up information is received from the stimulus by the decomposition of visual information into maps that represent the features, such as the color, size, form, and luminance of the package (Van der Lans et al., 2008). The bottom-up information is the imagery of the shelf, the information communicated by the individual package, and the saliency image. Information searching is, basically, a combination of top-down and bottom-up activities (Janiszewski, 1998).

The use of eye-tracking to study how consumers make in-store choices has made it possible to study issues such as visual attention during the choice process (Clement, 2007; Russo & Leclerc, 1994), the effects of time pressure during brand choice (Pieters & Warlop, 1999), how the saliency of a package can be understood in terms of its top-down and bottom-up qualities (Chandon, Hutchinson, Bradlow, & Young, 2007), and how the number of facings affects attention (Chandon, Hutchinson, Bradlow, & Young, 2009). Chandon and colleagues note that the center positions get more but shorter fixations (Chandon et al., 2009) which they interpret as ‘stepping stone’ fixations i.e. fixations during eye movements between locations that exceed the possible saccade length. They also note that products on the top shelves attract more attention than products on the middle or bottom shelves (Chandon et al., 2009). They do however qualify these findings as being artifacts of brand effects (Chandon et al., 2007) and preferences (Chandon et al., 2009).

For the purpose of investigating and describing how the eye tracking methodology can be used to evaluate packages we have conducted a series of three experiments an eye-tracking laboratory designed to facilitate studies on full-scale product shelves. In order to perform tests on full-scale shelves, an external eye-tracker is mounted on a specially developed foot-stand. Digital images were used to display the shelf stimuli. The underlying purpose is to understand how consumer read shelves during their shopping. In following sections these experiments are described.

**Experiment 1: Understanding the Inherent Properties of an Eye-tracking Laboratory**

All experiments are dependent on both context and equipment. The first experiment was therefore designed to learn more about the inherent properties of the eye-tracking lab. More specifically, it aimed to determine whether there are any specific viewing patterns, in terms of viewing order or package positioning that must be taken into account when studying digitally-displayed products. This study was designed to circumvent the issues of brand effects and preferences in order to ascertain the findings of (Chandon et al., 2007; Chandon et al., 2009).

**Method**

For the first experiment, 72 participants (42 female, 30 male, age: $M=34.97$, $SD=15.57$) were recruited in a supermarket during their regular shopping trips. Of these participants, 44.4% stated that they were the primary shopper in their household, 27.8% stated that their household’s shopping responsibilities were divided evenly, and 27.8% indicated that they were the secondary shopper in their household. The participants spent an average of
approximately 85€ ($SD = 70€) per week on shopping for their families (family size: $M = 2.85$, $SD = 1.27$).

In order to balance out the effects of specific brands and package designs, digital images of eight different shelves were created, containing everyday grocery categories such as coffee, cereals and crackers etc. The shelves were standardized, in that they contained three rows divided into five columns, which resulted in 15 products per image. All participants viewed all shelves, but the presentation order was balanced to cancel out any possible sequencing effects. This design made it possible to investigate the number of fixations that occurred before noticing a specific product, the percentage of respondents who noticed a specific product, and the mean observation length for each product.

Upon arrival, the participants were briefed about the objective of the study, the personal data handling policy, and their option to leave the trial at any point. The eye-tracking system was then calibrated for each individual. The trial began with a short demographic questionnaire and the viewing of a sample shelf, in order to familiarize the participants with the equipment and the stimuli used in the study. After this introduction, the participants were asked to look at the shelves as they appeared (viewing time: 12 seconds per shelf) and answer the questions that appeared after each shelf was shown. The participants spent an average of 22.5 minutes ($SD = 6.05$) completing the introduction and the test sequence that followed.

**Results**

The viewing patterns of the consumers were investigated while they looked at the digitally presented shelves. The initial analysis focused on the number of fixations a consumer had before looking at a specific product. An ANOVA was used to analyze the order in which consumers looked at the products. To summarize the results, an F-test $[F(4, 252) = 31.792, p<0.001]$ revealed that consumers generally start by looking at the middle section of the shelf and then move to either one of its two outer sections. With regard to vertical gaze patterns $[F(2, 126) = 42.465, p<0.001]$, consumers start by looking at the upper shelves and then shift their gaze downwards. The interaction effect between horizontal and vertical gaze movements is also statistically significant $[F(8, 504) = 9.943, p<0.001]$. In general, the consumers in the experiment started by looking at the middle of the shelf, then shifted their gaze to the top shelf, and then towards the outer parts of the shelf.

The number of participants who noticed a product on a specific part of the shelf was also analyzed. The ANOVA analysis revealed that most people notice the mid-position on the shelf, but that in general the products on the top shelf were noticed by most customers $[F(2, 126) = 6.176, p<0.01]$. Analysis of the noting pattern on the x-axis revealed that most participants noticed the products in the mid-position ($x = 3$), and that products positioned further out on the shelf $[F(4, 252) = 9.621, p<0.001]$ were noticed by fewer people. In this case, the interaction effect between the x- and y-axes is not significant, which suggests that the effect is similar on all shelves (see Figure 1).

**Figure 1: Noting – The percentage of participants who looked at the areas of interest**

*Note: X1:Y1 refers to the lower left corner of the shelf and X5:Y3 refers to the upper right.*

The mean observation length of the products in each position was also investigated. The ANOVA analysis revealed that the only significant difference concerned differences in observation length due to the position on the x-axis $[F(4, 64) = 8.564, p<0.001]$. The analysis revealed that the observation length was lowest in the middle of the shelf and highest at the sides of the shelf (see Figure 2).
The final part of the first experiment investigated the number of fixations on different parts of the shelf. Using ANOVA analysis, the results revealed a significant difference between shelves on the y-axis \[F(2, 126)=11.544, p<0.001\]. Consumers generally view the top shelves more than lower levels. With regard to the horizontal differences \[F(4, 252)=28.068, p<0.001\], the number of fixations was higher in the middle of the shelf than on positions further out on the shelf. The significant interaction effect \[F(8, 504)=5.845, p<0.001\] revealed that the mid-position on the shelf received the highest number of fixations and that there were more fixations further up on the shelf (see Figure 3).

Figure 2: The mean observation length of the areas of interest (i.e., the time between the gaze entering one area of interest and entering another).

Figure 3: Fixations – inverted heatmap based on the average number of fixations illustrating what the participants saw while looking at one of the shelves.

Experiment II: Evaluating the Findability of a Package

The second experiment sought to evaluate the findability of a package, that is, the likelihood of a package being found when looked for. The study was designed in order to solve a brand owners’ dilemma in which consumers repeatedly confused two of the company’s products that are intended for different applications. The brand owners’ design team created three new package designs for one of the two applications, all with the look
and feel of the brand but with the intention of placing extra emphasis on the specific application. The task was to investigate which of the three designs best facilitated the findability of the wanted product between the two alternatives. In order to do this, the investigation was set up to answer two research questions: ‘Which design is easiest to find?’ and ‘which design elements facilitates this process?’

Method

The majority of the 22 (11 female, 11 male) participants (age: 20-65) were recruited at a shopping centre during their regular shopping trips. All of the participants made at least half of their household’s purchases. All participants were regular users of the product and bought it at a supermarket on a regular basis. Forty-five percent of the participants said that they occasionally picked and bought the wrong package (that is, the product intended for use in the other application).

The study was designed as a straightforward response time experiment, whereby the participants were shown a series of pictures, each of which contained a set of four packages. Each set included three identical instances of the package containing the product intended for application A (used as distracters [D]), and one package containing the product intended for application B, which was the target [T] package (see Figure 4). The participants were asked to indicate, by clicking with a mouse, which of the packages was the target package, that is, the product intended for application B.

The target packages came in four varieties: the original and three new designs. In order to balance any effects of location and obtain a stable estimate of the response times, each of the four target packages was shown in all four possible locations, resulting in a total of 16 sets.

Figure 4: An example of one of the 16 sets used to investigate the findability of new designs [D=Distracter, Ti=Target]

Accordingly, each participant went through a trial containing 16 sets, and the mean response times to the four target packages were calculated. Additionally, eye-tracking data was used to determine which specific design elements facilitated or obstructed the identification process.

Results

Repeated measures ANOVA analysis showed a significant difference \( F(3,60) = 11.081, p<0.01 \) in response times among the four designs. Bonferroni-corrected pairwise comparisons showed that it took significantly less time to identify Design 3 than it did to identify the original and Designs 1 and 2 (see Figure 5).

Eye-tracking data was analyzed further in order to understand why this particular design was the easiest to find. The areas at which the participants had looked were examined in order to understand which package design elements they felt were informative. This revealed a pattern that made it possible to identify the design elements that were attracting attention but not necessarily conveying any choice-relevant information. Areas at which the participants had looked a lot were analyzed in order to understand which parts of the packages had been observed to a higher degree.

Figure 5: Mean response times (in seconds) to the original and the three new design alternatives
The results of this analysis revealed that Design 1 included a design element that had attracted particular attention. However, instead of conveying the intended message, which was the product's application, it simply indicated the weight of the product. Design 2 included a design element that highlighted the product's application, but this element was not being looked at. Design 3 included a design element that highlighted the application; the element was frequently looked at, which made this the easiest design to identify.

The second experiment, aimed at evaluating the findability of a package, provided some additional knowledge concerning how the eye-tracking methodology can be used to compare the role of different design elements in consumers' search process.

**Experiment III: Evaluating the Impact and Imagery of a Package**

An interesting aspect of package design concerns the issue of attention and awareness. In the second experiment, because there was no doubt about which package was the target package, all the participants selected the correct one. However, not all variations in package design are as apparent, which raises the questions of how small variations in the design can influence consumer attention and how such changes can be studied.

These questions were raised by a producer of packaging materials who wanted to learn more about the effects of various surface materials. Some of the effects of the different surface materials were quite subtle, so a physical replica of a supermarket shelf was built for the study, complete with the same type of lights used by stores, in order to make the display as similar to a supermarket as possible. It is worth noting that although this method is quite cumbersome in comparison to presenting digital images, it is still preferable if accuracy in visual characteristics is paramount. Two factors contribute to live shelf testing being more cumbersome. Firstly, the equipment must be set up so that it can accurately map the eye-tracking data on a digital image for subsequent replay and data analysis. Unlike eye-tracking based on digitally-displayed images, this method presents several accuracy pitfalls. Secondly, shelf or product images can be changed quickly during image-based studies, whereas changing the display during a live shelf test requires the test leaders to physically move the packages displayed on the shelf. This study was primarily designed to determine whether the participants would notice the differences in surface materials and, if so, how these differences would be perceived. The study also sought to determine how the differences, both perceived and non-perceived, would affect the viewing patterns of the participants.

**Method**

As with the previous experiments, all 34 participants (17 female, 17 male, age: 20–55) made at least half of their household’s purchases and all were recruited at a shopping centre during their regular shopping trips. Because the objective of the study was to investigate perceived differences between packages, participants were asked to view a number of displays, each of which contained three packages, and indicate whether they perceived any differences between the packages within each display. The participants who did perceive differences were asked to describe them. The displays were arranged in such a way that they contained one major difference and a number of minor differences. The analysis presented here is based on two of the displays: the one with the most divergent package on the left [DL] and one with the most divergent package on the right [DR].
Results

A 2 (Display[L/R]) * 3 (Position[Left/Middle/Right]) repeated measures ANOVA with Time to First Fixation as the dependent variable showed that there was a significant main effect for Display \[F(1,54) = 5.578, p<0.026\] and a significant main effect for Position \[F(2,54) = 7.892, \ p<0.001\]. The first of these results showed that content affects attention span, although this mechanism is not investigated here. The second is a result of the participants consistently starting in the middle. Of greater interest is the significant interaction effect \[F(2,54) = 8.337, \ p<0.001\], which shows that, after looking at the middle package, participants shifted their gaze towards the divergent package (see Figure 6).

Figure 6: Mean Time to First Fixation of the three positions during the two displays

![Figure 6](image)

Note: note how participants start in the middle and then shift their gaze towards the divergent package (on the right in Display DR and on the left in Display DL).

While looking at Display DR, all participants noticed the difference, whereas only 44% noticed the difference in Display DL. The next question is whether awareness of the difference is a prerequisite for attention capture. In order to investigate this, the response patterns of participants who noted the difference were compared with those who did not. A 2 * 3 mixed ANOVA with noting[Y/N] as a between-group variable and position[Left/Middle/Right] as an intra-group variable showed a significant main effect for Position \[F(2,54) = 3.930, \ p<0.026\], but no significant interaction effect (p>0.7) (see Figure 7).

Figure 7: Mean Time to First Fixation of the three positions during the DL display

![Figure 7](image)

Note: 0=participants who did not notice any difference and (1)=those who did. Both groups begin in the middle and then shift their gaze towards the left.

The third experiment, which was aimed at understanding how small variations in the design influence consumer attention, provided additional knowledge concerning the usefulness of the eye-tracking methodology. Even when consumers are not aware of design differences, they are able to identify a divergent package. Based on the results of the three experiments,
the study attempts to evaluate the usefulness of the eye-tracking methodology in the development of packages.

**Discussion and Implications for Future Research**

The objective of the three experiments was to building an understanding of the process a consumer uses when reading shelves in a retail setting. The impetus behind this approach is the important role of the visual design of a package while choosing FMCGs (Silayoi & Speece, 2007; Underwood & Klein, 2002).

The results of experiment 1 showed three properties of lab shelf testing that must be taken into account during the design of shelf studies. Firstly, participants’ initial fixations tend to be in the middle of the display. Secondly, in the absence of any particular instructions, participants’ gazes shift upwards and outwards and then turn into a more or less random pattern. Thirdly, gaze data shows that consumers have more fixations in the center of the display and longer observations around the sides. These findings corroborates the results of (Chandon et al., 2007; Chandon et al., 2009) as it is not biased by brand effects or preferences. The major implication of this result relates to positioning. Fortunately, the problem can be solved, and utilized, by adhering to the principles of experimental design. The biases of the observed gaze pattern and gaze distribution can be corrected by balancing the experimental design for position effects. In addition, observing where the gaze shifts to from the initial middle fixation and changing the adjacent packages will make it possible to investigate the impact of these changes on the viewer’s visual attention.

The first proposed method was designed in order to evaluate four new design alternatives in terms of the findability of a package, that is, the possibility of a certain design being identified. Consumers generally spend a maximum of six seconds selecting an item (Economist, 2005), so the ability of a product to catch their attention will facilitate this decision-making process (Silayoi & Speece, 2004). Findability is therefore a key aspect during the choice process. In this case, the objective is to solve the consumer’s problems of identifying and selecting their preferred product within a brand owner’s range of products. Overall, the results of this study clearly show which of the proposed design alternatives is the easiest to find and which elements of the design facilitate findability. Two points regarding the design of the study are worth extra consideration. Firstly, the experimental design, with all target stimuli appearing in all positions, makes it possible to draw valid conclusions about the effects of package design without the data being confounded by the effects of target positions. Secondly, the combination of a performance measurement, such as a response time, and process data, such as gaze data, provides clear indications of which package is the easiest to find and why. The methodology proposed here therefore makes it possible to evaluate novel packaging designs compared to other packaging that might be competing for the consumer’s attention.

The second method was designed to investigate the impact of a proposed minor change to the appearance of a package. The objective was to determine whether the participants noticed the change and, if they did, what they thought of it. In order to make the test as ecologically valid as possible, the packages were shown on a real shelf in the laboratory. Participants were shown three packages, two that were identical and one with a divergent design. The participants were asked to look at the three packages and identify any differences. Most of the participants shifted their gaze from the center package to the position of the new design, regardless of whether they reported observing any differences. This indicates that visual variation did not affect conscious processing, but that it did affect the unconscious attention process. The finding that impact drives attention is in line with previous research on visual saliency (Itti & Koch, 2001), yet the extent to which attention was guided by the stimuli’s visual saliency or by its relevance to task completion is unclear. This question should be addressed by further research as it points to two different areas of packaging development; one emphasizing visibility vis-à-vis other products on the shelf and one emphasizing the content and information on the package itself. Although the principle of live shelf testing in the laboratory worked out satisfactorily, and the participants had no difficulty describing their subjectively perceived imagery of the packages, the packages included too few design elements to determine, with certainty, the parts of the packages that conveyed specific imagery.
References
