The Dating Game – Understanding Expiration Phrase-Date Salience Using Eye Tracking Technology

Talia Katz\textsuperscript{a}, Aishwarya Badiger\textsuperscript{a}, Christopher Simons\textsuperscript{a}, Brian Roe\textsuperscript{b}

\textsuperscript{a}Department of Food Science and Technology, The Ohio State University
\textsuperscript{b}Department of Agriculture Environment and Development Economics, The Ohio State University
1.1 Abstract

Expiration dates on food products have been known to promote food waste in consumer households. Current policies on date labeling target standardizing the phrases without any efforts to expand date horizons. The current research in this field is unclear regarding whether the date (Nov 29) or the phrase (e.g. sell by) has the greatest impact on food discard. Eye tracking technology is a useful tool which can be used for visual analysis. This implicit technique allows researchers to see beyond just the explicit feedback that participants provide and see truly what they are looking at. Eye-tracking was used to analyze how consumers look at milk expiration labels and what information ultimately encourages them to keep or discard foods. The study was conducted using a mixed design with 68 participants. Panelists wore eye-tracking glasses throughout the study and saw multiple milk carton images on the screen featuring different dates. Along with the images, they were given physical milk samples to smell which they were told were from the milk carton on the screen. Data was processed in Tobii Pro Lab software and a Mixed ANOVA model was used for statistical analyses. The results indicated that viewers tend to fixate on the date more than the phrasing (p<0.001) when making their discard decisions, both in the presence and in the absence of a real shelf-life indicator. On average, 50% of the panelists did not look at the phrase at all. Panelists tend to fixate more on the date when a “use by” label was present as opposed to “best if used by” or “sell by” label (p=0.037). Since the label phrase is rarely looked at when making discard decisions, standardizing the phrases is unlikely to reduce food discard by itself. The results from this study inform current policy development surrounding date labeling with the potential to influence food discard and consequently food waste.

1.2 Introduction
In growing discussions to create a more sustainable planet, it has been found that food sustainability has been of increased conversation to combat the large amounts of food waste generated. Food has a substantial impact on the economy regarding producing and consumption. Aside from the economy, food is an important aspect of culture, health, and wellbeing and is a key driving force in society. The food sector causes a third of all greenhouse emissions in the EU (Witzel, Hooge, Amani, Larsen, Oostindjer 2015). While food production can have a significant impact on global sustainability, the majority food waste in developed nations is at the consumer level, which is considered the final stage of the food supply chain (Schanes, Dobernig, Gözet 2018). Consumer perception of freshness and food safety is a leading cause in the production of food waste.

A leading factor in the perception of freshness and safety is the use of date and phrase labeling on products. Since there is not a federally or state-wide regulation in the United States regarding phrase labeling, consumers are faced with a wide array of phrasing (Leib, et al 2016) eg., best by, use by, sell by etc. Approximately 84% of consumers discard products based on the date label, perceiving that it is an indicator of food safety, whereas it is used by manufacturers as a prediction of how long they think the item will retain its peak quality (Leib, et al 2016). In addition to date label confusion, phrase labeling provides another aspect of confusion that can be detrimental to food waste. A 2016 study found immense diversity among the consumer interpretation of labels finding that “best if used by”, “best by”, and “freshest by” serve as indicators of food quality while “expires on” and “Use by” were interpreted by consumers as being indicators of food safety (Leib, et al 2016).

To minimize phrase confusion a 2018 initiative led by the Food Marketing Institute (FMI) and the Grocery Manufacturers Association (GMA) has begun streamlining “Best If Used By” as
a standardized phrasing on products. However, further studies have found that the standardization of the phrase labeling does not have a significant effect on discard decisions of consumers (Bender 2019). The relative importance to consumers of date versus phrase labeling can be studied with eye tracking technologies. Video-based eye trackers can be used to look at gaze directions, patterns, fixations, and many other metrics which are able to be computed from the movement of participants eyes (Carter, Luke 2020). The present study uses eye-tracking technology to determine the influential portion of food labeling when looking at consumers’ discard decisions of various qualities of milk. Additionally, this study looks at the interaction of the various phrase labeling, “Best If Use By”, “Sell By”, and “Use By”, as well as age and quality on the visual and discard patterns of consumers. Discard patterns are also noted and studied when a novel Real Shelf-life indicator (RSI) is present, to see its impact on both discard decisions and panelists eye patterns.

1.3 Materials and Methods

1.3.1 Experimental Design

Panelists (n=70) were screened through The Ohio State University Sensory Evaluation Center database based on their age, 18+, as well as their milk purchase and consumption patterns. Each of the seventy panelists were asked to evaluate different samples of milk served in 1 oz jars which each correlated to the image of a milk carton in the survey. Panelists evaluated various milk samples as well as images which contained different date label phrasing. Panelists answered how they would manage the product with options of “keep”, “discard”, or “unsure”.

Label phrases varied between “Use By”, “Sell By”, and “Best If Used By” and were sometimes used with real-time shelf-life indicators (RSI) as well as a no label treatment. For each panelist, four different levels/dates were generated for the images (Table 1). Levels varied
from baby (B), young (Y), medium (M), and old (O). The baby milk portrayed a date label with 6 days remaining from the present. The young milk had two, the medium was 1 day past the date, and the old milk was 7 days past the date. For each date label received, two different milk quality samples were evaluated, one of good quality and the other of poor quality. This was a mixed-design study with phrase as a between subject variable and quality, product age, and label area of interest (phrase or date) serving as within subject variables. Each panelist received three different flights of samples, with the first two flights containing 8 milk samples. For the first flight panelists were asked to evaluate the samples while viewing images containing a phrase and date label. The next flight contained images with phrase, date, and RSI labels. The RSI label contained information that correlated to the quality of the milk, hence poor quality showed zero days remaining and good quality reflected the number of days left that the date label had. The final flight contained four milk samples and four images that portrayed no date labels.

1.3.2 Milk Samples

Two percent milk was collected and received from Tamarack Farms Dairy. A total of nineteen gallons and thirty-eight half-gallon containers were collected. From the milk collected, 9.5 gallons and nineteen half-gallon containers were immediately frozen to retain good quality. The other half of the milk was stored in chest freezers at 7°C until a poor quality was achieved. The milk was checked every third day for the first 10 days and then daily for the last 9 days for a total of 19 days. Milk was evaluated based on appearance, smell and given an odor score, with ten being good quality and one being horrible quality. In the early stages, 3 cartons from each chest were selected at random for evaluation, but once changes in quality were noticed each carton was evaluated. Once the poor quality was achieved for each container, they were immediately frozen. Upon preparation for the study, milk was removed from the freezer and
thawed. Once it thawed, the 20 mL of milk was filled into 1 oz glass containers that was sealed with airtight lids. The jars were then frozen and removed from the freezer about 36 hours before the day for which they were needed. The good and poor-quality samples were visually identical but differed in smell.

<table>
<thead>
<tr>
<th>Number of Panelists</th>
<th>Flight Number</th>
<th>Label Phrase</th>
<th>Date/RSI days remaining</th>
<th>Milk Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1</td>
<td>Sell By</td>
<td>+6,+2,-1,-7</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sell By + RSI</td>
<td>+6/+6,+2/+2,-1/-1,-7/-7</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>No label</td>
<td>- (two reps)</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No label</td>
<td>- (two reps)</td>
<td>Marginal</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>Best if used by</td>
<td>+6,+2,-1,-7</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Best if used by</td>
<td>+6,+2,-1,-7</td>
<td>Marginal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No label</td>
<td>(two reps)</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No label</td>
<td>(two reps)</td>
<td>Marginal</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>Use By</td>
<td>+6,+2,-1,-7</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Use By + RSI</td>
<td>+6/+6,+2/+2,-1/-1,-7/-7</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No label</td>
<td>(two reps)</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No label</td>
<td>(two reps)</td>
<td>Marginal</td>
</tr>
</tbody>
</table>

Table 1. Represents experimental design used for the study showing the various treatments such as labels, dates, and qualities received by panelists.

1.3.4 Milk Carton Images

Half gallon milk cartons filled two thirds of the way with milk were used. Acetone was used to remove the preceding date label and was replaced using a handheld date label printer. Date labels were created according to Table 1, using the “Sell By”, “Best if used by”, and “Use by” phrasing. Dates were assigned based on the day of the study. Stickers of RSI labels (Avery Products Corporation, Brea, CA, United States) was used to mimic real RSI labels. The stickers resembled the information portrayed by the date label for the good quality milk and displayed either 2 or 0 days remaining for the marginal/poor quality milk. Three-digit blinding code
stickers were applied to the cartons above the date and or the RSI label. The cartons were tilted against a black background with the help of a mount. A Canon EOS Rebel T6 camera with 20 mm lens, f/9 aperture setting, 1/50sec shutter speed and 200 ISO was used to collect images of each carton.

1.3.5 Survey Measures

Qualtrics surveys were used for both the panelist selection process in addition to the collection of data for the study. Initially, individuals were screened by their age, consumption of milk, as well as the frequency of purchase of milk. Once selected for the study, participants selected 1-hour time slots to participate in the study at the Parker Food Science Building, room 054. During the study, the survey consisted of four different sections, with three of them set up the same way. Participants were shown a screen with an image of a milk carton with a 3-digit blinding code and were asked to locate the same sample from the tray provided to them. The survey asked the participants to smell and evaluate the milk while simultaneously referring to the image of the carton provided in the survey. The panelist was required to remain on this page for 30 seconds prior to an arrow being available to them to move onto the next page. In the following page, the survey asked the panelist whether they would keep, discard, or were unsure about the product. If keep or discard were selected, then they were prompted to fill in how many additional days they would keep the product for. This was repeated for a total of 20 questions. The final section of the survey asked the panelist more about their milk consumption patterns. They were asked what characteristic or quality of milk they deemed important when making decisions about milk products are home. Additionally, information regarding demographics was also collected. Ultimately, the survey used tracked their responses to the questions in addition to the time spent on each page and each question.
1.3.5 Eye Tracking Technology and Data Extraction

Tobii Pro Glasses 2 were used for each panelist during the study. The Tobii Pro software (Tobii Technology, Stockholm, Sweden) was used during the study to collect data for each panelist which was ultimately converted into the desired metrics such as fixation duration, visit duration, visit count, fixation count. Images from the participant surveys were uploaded to the software and uniform areas of interest were drawn using the software on every image used. The AOIs were separated into date, phrase, and real shelf-life indicator (RSI).

Assistant mapping features were used to take participant data and plot the information onto the desired pictures and AOIs.

1.3.6 Data Analysis

Data were analyzed using SPSS statistics software. Univariate mixed ANOVAs were run to understand the effect of AOI, phrase, quality, and product age on each of the eye tracking metrics when an RSI label was and was not present. Additionally, ANOVA tests were run to understand the effect of date and phrase labeling compared to the RSI label. Post-hoc Tukey tests were used to compare means for the significant terms. Pairwise comparisons were made for the interaction terms with Bonferroni p-value corrections.

1.4 Results and Discussion

1.4.1 Phrase and Date AOI

Using Tobii Pro Software, creation of Areas of Interest (AOI’s), and assistant mapping features data were able to be statistically compared using SPSS statistics software. In Figure 1, representative AOI regions were drawn on milk cartons distinguishing between the phrase and
date regions of the cartons. Dimensions for the two regions were kept constant for all cartons and for all heat maps generated. Figure 1 also illustrates the heat map generation process where the red intensity spots signal fixation attentions which lasted up to two seconds total. This figure ultimately shows that the red areas with the highest amount of fixation were located on the date regions rather than the phrase AOI. It was ultimately found that 50% of panelists fixations did not include the phrase AOI, signifying that in some cases the phrases was not even looked at. In addition, it was found that the date label was found to be what rapidly attracted participants based on the Time to First Fixation metric. This ultimately found that the date region of the label is what attracts the consumers attention and gaze first, holding the most significant meaning for the consumer. In table 2, the figure represents the mixed ANOVA tests for the comparison of phrase, product age, quality, AOI, and phrase AOI for the Total Fixation Duration metric, it was found the phrase type as well as the product age were deemed nonsignificant terms. It can be concluded that the phrase and product age were not influential in the discard decisions made by the participants.

When analyzing the specific significance of the date and phrase AOI for the Fixation
and Visit Duration, there was a significant difference between both metrics for date and phrase AOI. Figure 3 represents the Total Duration of fixation and Visit Durations for both date and phrase AOI, with a significance level of $p<0.001$ it was found to be a significant difference in lengths for the two metrics. There was a significantly longer duration for the date AOI compared to phrase AOI. Likewise, similar metrics were found across all the various phrase types. Figure 4 represents the metrics found when comparing the fixation durations for the various phrase times. At a $p$ value of 0.001, there was a significantly larger fixation duration for the date AOI versus the phrase AOI. It was found however, that the date region was fixated on more in the presence of the “Use By” label compared to the date AOI for the “Best If Used By” phrase ($p=0.034$). While it was found for some metrics, this wasn’t found across all metrics.

1.4.2 Phase and Date Significance with RSI Label
When understanding the impact, the real shelf-life indicator had on the visual fixation and visit duration, it was found that in the presence of the RSI label, the date region held a higher significance relative to the phrase AOI. Additionally, Figure 5, represents the Total Fixation Duration metrics for the date and phrase AOI under the influence of the RSI, revealing that there is still a strong significance of the AOI as well as the interaction between the phrase and AOI (0.013) at the significance level of (p<0.05). The estimated marginal mean for the date AOI was 0.810, and for the phrase AOI was 0.236. Additionally, when comparing different phrase types under these conditions, significant differences exist in the pairwise comparisons between the “best if use by” and “sell by” (0.021) and “best if use by” and “use by” (0.043) at the significance level of p<0.05. However, when comparing the “sell by” and “use by” phrases, there was seen to be no significant difference between the two phrases (p=0.064). Additionally in the presence of the RSI, the quality of the milk did not have an impact on either AOI regions. In addition to fixation duration, total visit duration also provided similar information finding a significance between the phrase and date AOI (p<0.001). The estimated marginal means for the phrase AOI was 0.175 seconds compared to 0.737 seconds for the date AOI. For the various phrase types however, no significant differences were found between the different phrase AOIs.

1.4.2 Date and Phrase AOI comparison to RSI AOI
In the presence of the RSI label, the RSI AOI was found to be fixated on more. In the presence of the RSI, the RSI label was fixated on 3.8 times more than the other areas of interest. A significant difference (p<0.001) was found for the Total Fixation Duration showing a significant differentiation between participants viewing the RSI label and the combined phrase and date AOI. Likewise, the Total Visit Counts for the two AOIs found significant differences between the AOI and Date/Phrase AOI with increased visit counts to the AOI region (p<0.001). When looking at the affects which the presence of AOI had on the visits and fixations of the date and phrase AOI, the results were consistent with previous findings such that the date was steadily visited and fixated at more than the phrase AOI was (p<0.001).

1.5 Conclusions

The study found that the date region is what rapidly attracted the participants attention and was more attractive than the phrase region. It was found that regardless of the various phrase types, “Best If Used By”, “Sell By”, and “Use By”, the different phrase types did not have an impact on the Visit Counts, Visit Durations of the phrase AOI and that the date was looked at the most regardless of the phrasing. However, the Fixation Duration metric did find that the date AOI was fixated more in the presence of the “Use By” label relative to the “Best If Use By”, however this finding was not seen across other metrics. Additionally, this study found that other variables such as age and milk quality were non-significant factors.

With the presence of an additional AOI, the real shelf-life indicator (RSI), similar results were also found. When the RSI AOI was present, it was still found that the date AOI was more
significant than the phrase AOI, with a greater number of fixations, visits, and visit durations. However, when comparing the various phrase types, it was found that significant difference existed between the “Best If Used By” and “Sell By” as well as “Best If Use By” and “Use By” phrases regarding total fixation duration metric. However, other metrics did not display the same significant differences between phrase variations. The quality and the age of the milk did not however have impacts on the date and phrase AOI with the RSI present.

When comparing the RSI to the combined date and phrase AOI, the RSI was looked at significantly more than the date-phrase AOI. Additionally, under these conditions, the date AOI continued to receive more visits and fixations than the phrase AOI. This could be so because the RSI is a novel food label and consumers in the United States are not used to having these labels present. Since they are something new, it could cause consumers to be attracted to them more and spend more time looking at this label to understand what they are and their uses.

Going forward, this study is useful guidance for the current attempts to standardize food labels. Policies so far have been focused on creating a standardized label, however this study suggests that the standardization of a food label isn’t what matters to the consumer and that the date is more important to a consumer than the phrase and is often the cause of discard decisions. This study ultimately would suggest instead of creating a standardize label, it would be more useful to implement a label, like a real shelf-life indicator, that can give the consumer a better idea of the overall quality of the product which the date is unable to judge accurately.

1.6 References


