8-2015

# PREFERENCE OF CONSUMERS TOWARD NON-DISTORTED GRAPHICS ON FULL-BODY SHRINK SLEEVE LABELS

Nathan Bailey
*Clemson University*, nbailey@g.clemson.edu

# PREFERENCE OF CONSUMERS TOWARD NON-DISTORTED GRAPHICS ON FULL-BODY SHRINK SLEEVE LABELS

---

A Thesis
Presented to
the Graduate School of
Clemson University

---

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Packaging Science

---

by
Nathan Bailey
August 2015

---

Accepted by:
Dr. Rupert Andrew Hurley, Committee Chair
Dr. Duncan Darby
Dr. Chip Tonkin

ABSTRACT

Past research has indicated that shrink sleeves can lead to higher product trial rates, better long-term sales, and greater likelihood of brand loyalty. A pilot study and a primary experiment were conducted to investigate the significance of distortion and the ability of the Human Visual System (HVS) to recognize it in packaging design. Distortion works primarily in shrink film on which an image is printed, so these studies dealt only with reductive distortion.

The pilot study aimed to identify the absolute threshold, or Just-Noticeable-Difference (JND) for a change from no stimulus, for simple polygons. The primary experiment focused on graphic distortion in full body shrink sleeves (FBSS). Treatments presented each of the stimulus levels, along with a control for comparison, using a 2-AFC (Alternative Forced-Choice) Method. This study used a mixed 2 (labels) x 3 (bottles) x 5 (distortion increments) model, effectively 30 treatments in a Randomized Complete Block Design (RCBD). Label was a between-subject variable while Bottle Shape and Distortion Percentage were within-subject variables.

Data indicated that distortion has a significant effect at 100%, but that there is not a threshold at which consumers are guaranteed to perceive graphic distortion on FBSS. Men detected distortion better than women. Participants who said distortion would prevent a purchase decision had the same tolerance as those who reported that it would not. Bottle shape may only impact consumer acceptance of distortion with some percentages of distortion. Familiarity with a brand name label may increase consumer tolerance and acceptance of distortion, and may do so more on some bottle shapes.

DEDICATION

To my parents, Mark and Jill Bailey, my brother, Jonathan, and my sister, Jennifer, for their ceaseless patience, support, and encouragement.

And to my friends Chris Gottilla, Nathan Schneider, and David Rauch. Iron sharpens Iron.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Rupert Andrew Hurley, and my committee members, Dr. Duncan Darby and Dr. Chip Tonkin, for their guidance and support.

Thank you to Ben Sharp, Dr. James Rieck, Dr. Patrick Gerard, and Dr. William Bridges for the immense support with the statistics.

I also greatly appreciate the input from Doug Becker at Fort Dearborn.

Thank you also to my fellow graduate students in the packaging design program who have set the pace and helped me throughout my research, including Toni Gomes, Joanna Fischer, Daniel Hutcherson, Rachel Randall, Katie Thackston, Erin Snyder, Wilson Sansbury, Felipe Fernandes, David Cottrell and David Stone.

Finally, I would like to thank Harris A. Smith and all those involved with the Sonoco Institute of Packaging Design and Graphics. Thank you, Kim, for letting me keep the keys for so long, and Bobby for bringing your sense of humor to everything.

This research would not have been possible without the help of these individuals.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

Table Page

List of Tables (Continued)

Table Page

## LIST OF FIGURES

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

# CHAPTER ONE

# INTRODUCTION

Past research has indicated that shrink sleeves can lead to higher product trial rates, better long-term sales, and greater likelihood of brand loyalty. Most full body shrink sleeve (FBSS) labels reduce in circumference by 40% to 65% when applied to the bottle. A great deal of work and many adjustable parameters are required to achieve adequate, repeatable shrink. It is difficult to produce and apply the label appropriately. A change in graphics, or distortion, occurs even in a successful production run. This research was conducted to investigate the significance of distortion and the ability of the Human Visual System (HVS) to recognize it in packaging design. Distortion here works primarily in shrink film on which an image is printed, so these studies dealt only with reductive distortion. From a consumer perspective, how much room for error is there in graphics? How much time, effort, money should be spent on perfecting graphics on FBSS labels? What other variables, like bottle shape and familiarity with graphics, could impact consumer perception of distortion?

Gustav Theodor Fechner's research during the late nineteenth century, which has been repeated numerous times, concluded that the rectangular proportions most used and preferred are those of the Golden Ratio, 1:1.618. (Elam, 2001) He did this by sampling every day objects and surveying ratio preferences along a spectrum. This information leads to some questions: If people gravitate toward this ideal, could they successfully accomplish the inverse, identifying among many objects one that is inconsistent with the ideal? To what degree are people capable of detecting when an ideal image has been lost through distortion? Basics of art and design are universal, according to Lidwell, Holden, & Butler (2010), and surely play a role in distorted images, as they do between normal images. Research was needed on the fundamentals of vision, perception, image assessment, graphic manipulation, and sensory experimental methods.

CHAPTER TWO

REVIEW OF LITERATURE

### ***Products, Packaging, and Appearances***

People unconsciously make little to no distinction between a product, the packaging, and their associations. Louis Cheskin of Ukraine investigated how design elements impacted people's perceptions of value, appeal and relevance in the mid 20th century. He observed that people’s perceptions of products and services were directly related to aesthetic design, and most people could not resist transferring their feelings towards the packaging to the product itself. He named this relationship sensation transference. (Cheskin, 1951) His research had reverberating influence in the development of iconic brands. This extends to the choice of color on Ronald McDonald, the addition of the spoon on Betty Crocker labels, the market research behind the Ford Mustang, and the creation of the Marlboro Man and the Gerber Baby. Most famously, his work led to the use of yellow coloring and a foil wrapper with introduction of the first mainstream margarine, Imperial Brand. He advocated that directly asking customers what they think of a package design is not a useful way to measure effectiveness. Surveys and polls can’t measure unconscious reactions; what consumers do, not what they say, is what matters. Most people who claim advertising and packaging don't affect them tend to buy widely advertised products. Advertising aside, sales and market shares can be gained by eye-catching and self-promoting packaging. (Godin, 2003) Excellent packaging influences as much as two-thirds of purchasing decisions by consumers. (Rettie &

Brewer, 2000) In hypermarkets, which are combination supermarket and department stores or wholesale warehouse clubs, there can be nearly 40,000 kinds of packages, and consumers can make 70% of the purchasing decisions at the point of purchase (Seal, 2009).

Packaging is hardly considered in isolation, and so is always geared toward relative differentiation from and preference over competition. (Young, 2006) Like advertising, packaging that stands apart is best since consumers prefer what they can easily identify. (Gobé, 2001). The purpose of packaging is to protect and label products, catch consumer attention from the shelf, set products apart from competitors by exhibiting advantages, and compel purchases (Doyle, 1996).

Shape, size, color, pattern, texture, character, brand, and smell must be considered for packaging to be effective, according to past studies by Kotler (1997) and Smith & Taylor (2004). Before label information can be processed, attention must be attracted by novel perceptual stimuli, like package form or in-store displays, which can interrupt shopping routines of consumers. Product attention and differentiation consists of four aspects, seen in Figure 1: properties of commodity (product and brand), label design (language, image, and color), package or bottle shape (height, width, volume, shape of bottle neck, shoulder, mouth, handle, and color of the bottle and cap), and interference factors (degree of memory, cognition, and attention). Properties of commodity have the greatest influence, bottle shape design second, and label design has lesser influence in product selection process. When a consumer has a preference for certain brands, they can quickly scan the shelf, pick out their favorite package by its curved bottle, and place it in their shopping cart (Garber, 1995). Any aspect of package design can have a strategic impact on consumers. (Orth & Malkewitz, 2008)

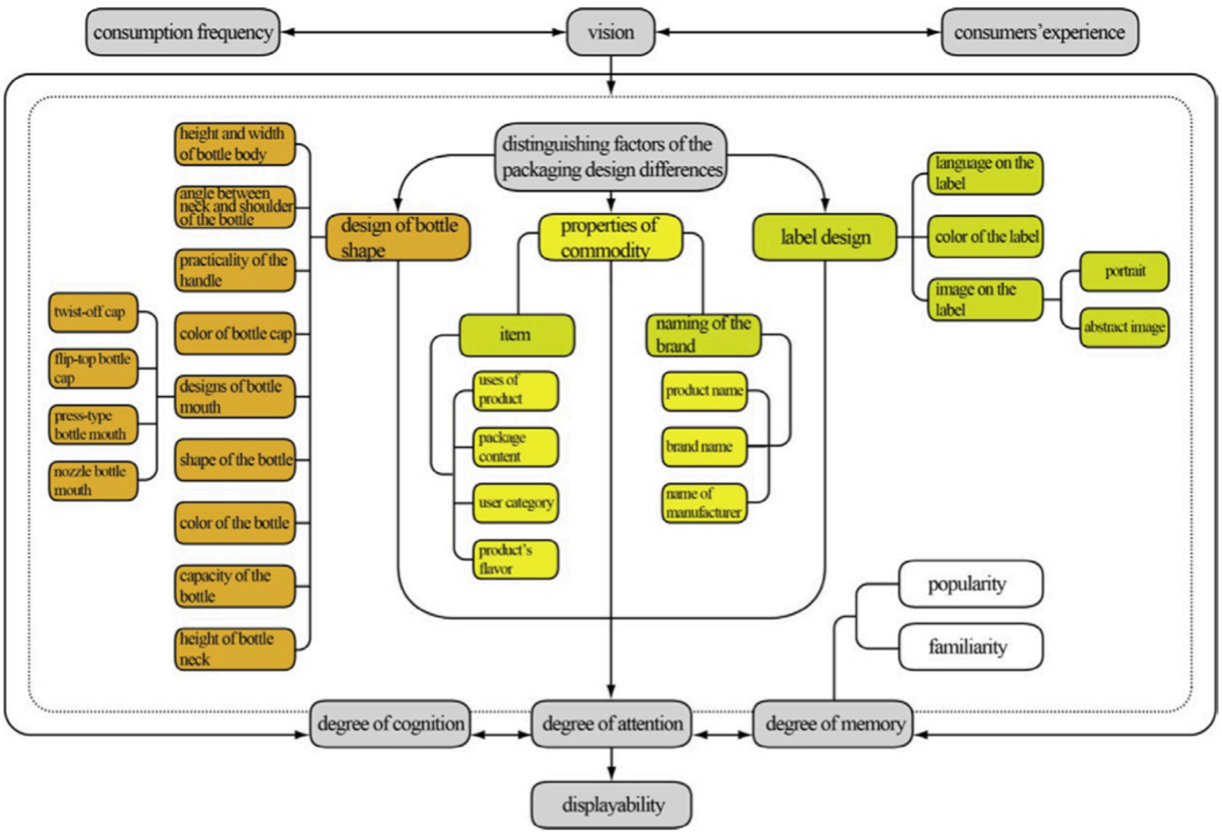


Figure 1. Conceptual model of distinguishing package differentiation. (Chou & Wang, 2012)

Appearances, especially shape and size, can consistently impact perception and consumption of a product. Research has shown that both students and bartenders over pour, and do so more into glasses that are short and wide than glasses that are tall and slender. (Wansink & Van Ittersum, 2005) Practice can reduce the tendency to over pour for only tall, slender glasses. Despite an average of six years of experience even bartenders tend to pour 20.5% more into short, wide glasses. Paying careful attention reduces but does not eliminate the effect. Because people generally consume most (about 92%) of what they have served themselves, the issue of perception, pouring accuracy, and consumption is relevant to law makers and enforcers, doctors, responsible consumers, and packaging designers. With labels that can conform to bottle shapes, appearances can have even further impact on consumers.

The shape of packaging helps define brand characteristics and deliver messages, both literal and subliminal, about the contents. It can establish visual and emotional connections with products. This is evidenced by package shape, and bottle forms especially, being trademarked and registered. According to Bar & Neta (2006, 2007) there is a general preference for contoured objects that are rounded as opposed to angular and a preference for upward shape orientation. The concept of typicality is that images or objects will be more recognizable and memorable the better they fit into expected categories. In research by Westerman et al. (2013), consumer preferences could not be accounted for by design typicality, but simple congruence between graphic and product forms seemed advantageous according to ratings of purchase likelihood.

Congruence can be maintained while graphics or materials are altered to enhance and invigorate a brand image. Coca-Cola is a prime example. It was packaged in a glass bottle developed in 1915, expanded to plastic in 1978, to aluminum in 2005, and eventually utilized its now iconic contoured form in all mediums. (Contour Bottle History, 2015)

Another is Budweiser, which believed so much in the influence of packaging appearance that it redesigned the aluminum beer can in an effort to reflect its iconic bowtie logo, which debuted in a national advertising campaign in 1956. The new can is a natural progression from the new label introduced in 2011 that emphasized the bowtie. Pat McGauley, vice president of innovation for Anheuser-Busch, said “We explored various shapes that would be distinguishable in the marketplace, but also viable from an engineering standpoint. The world’s most iconic beer brand deserves the world’s most unique and innovative can. I think we have it here. This can is certainly a conversation starter: eye-catching, easy-to-grip, trendy and – according to our research – very appealing to young adults. It’s a beer can like no other.” (Budweiser's 'Bowtie Shape' Can, 2013) This was an innovation of considerable investment led by the belief that congruence in packaging material, form, and graphics has a profound impact on consumers.

A third relevant example is Sobe Lifewater, seen in Figure 2. An AC Nielsen Consumer Study in 2010 found that in in the first year after switching from roll-fed label to full body shrink sleeve (FBSS) label, average monthly sales of Sobe Lifewater grew 103% while the flavored water market declined by 8%. The second year, average monthly sales grew 45% while the flavored water market grew only 2%. They concluded that “shrink sleeves, especially combined with high-contour bottles, possess a superior overall appearance, command more shelf attention, and create stronger emotional connections for consumers. As a result, higher product trial rates, long-term sales, and brand loyalty are likely.”

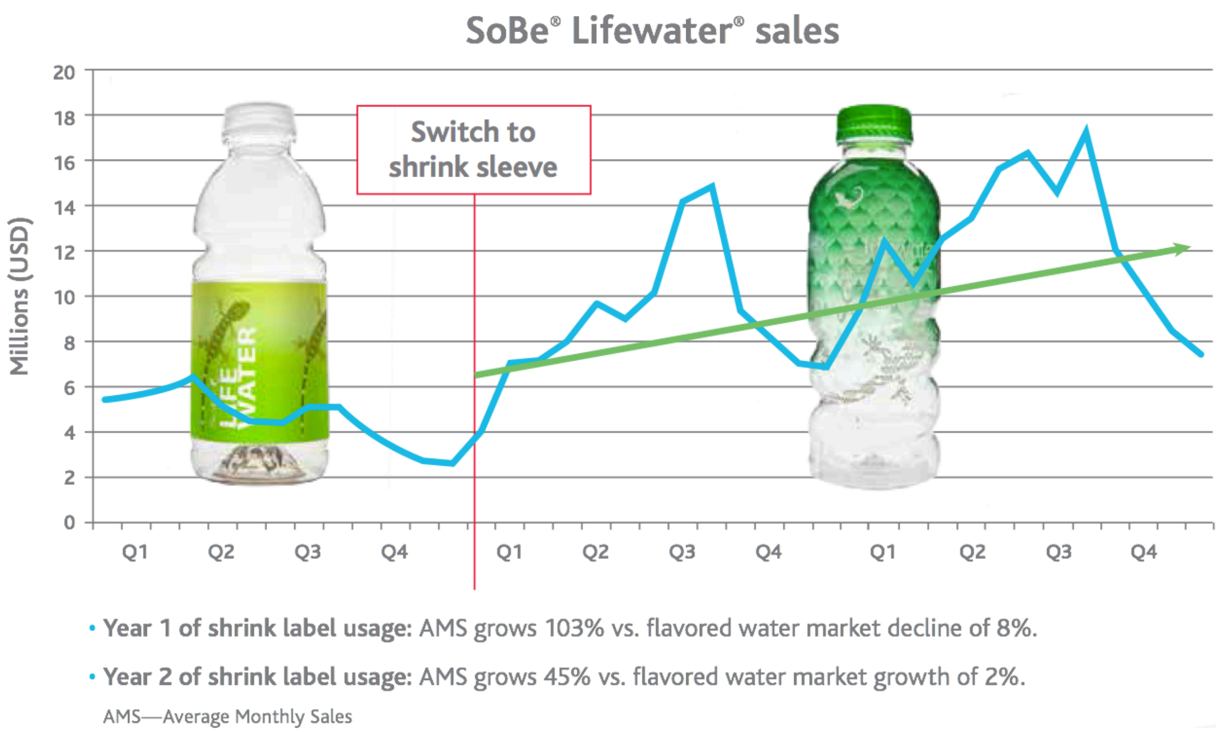


Figure 2. Increased sales of Sobe Lifewater after introduction of FBSS (AC Nielsen Study, 2010)

With the average shopper spending five to seven seconds scanning a label on the shelf, it must be effective in communicating its message. (Roncarelli & Ellicott, 2010) Labels that are compelling yet clear can stand apart from competition. Minimalist design has grown in popularity as it reduces visual stimulation in a busy shelving environment. Color, the first feature about a package that consumers notice, can impact mood and influences perceptions of size, quality, value, and flavor. (Roncarelli & Ellicott, 2010) Thus, it is an important motivation for purchase decisions. Bright colors can appear warmer, larger, and appear to move or radiate outwards. Basic process colors can seem cheap while subtler colors convey quality and refinement. Messages grouped around two or three benefit words or phrases are best, according to eye-tracking studies. More and more, beverages are viewed as specialty or gourmet food items. This is due to their increasing ability to emphasize nutrition and fashion. (Roncarelli & Ellicott, 2010)

Bottle making injection-molding techniques encourage designs that are rounder in nature, avoiding corners and angles. (Westerman et al., 2012) Packaging designers now must think like sculptors, considering how packaging must work in every way to appeal to the emotions, logic, and senses of consumers. Bottle shape has proven to differentiate a product on the shelf, as it draws people's attention to the product. The success of packaging design is measured by the ROI, or return on investment. ROI measures performance through qualities like volume of consumer recalls, visibility on shelf, appeal to aesthetics, differentiation from competition, and personal relevance. In addition to traditional advertising, improving packaging design has the ability to increase sales online and through word-of-mouth campaigns. Focus groups are useful in learning trends and strategic concepts but are less helpful in testing retail behaviors because group dynamics can sway preferences. When it comes to testing packaging design, eye-tracking can reveal consumer response prior to market launch.

***Development of Shrink Sleeves***

In 1988, the Society of the Plastics Industry created the plastics numbering system, which helps recyclers sort major types of plastics. (Roncarelli & Ellicott, 2010) #1 is PET, polyethylene terephthalate, the most commonly used plastic material, seen in bottles and food packaging. It is light, clear, and durable. #2 is HDPE, high-density polyethylene. It is less clear but more durable, and is used for larger or thicker containers. #3, PVC or polyvinyl chloride, is extremely tough and weathers well. It is rarely recycled because of these qualities and its use in disposable packaging has been greatly reduced in recent years. It is made into piping, medical equipment, and construction materials. #4 LDPE, low-density polyethylene, is both tough and flexible and used for all types of plastic bags. #5 is PP, or polypropylene. It is frequently used for containers of hot liquids due to a high melting point. #6, PS or polystyrene, is used in trays, cases, and foam packaging.

Primary packaging is manufactured from PVC, PP, PE, PET, and PEN (polyethylene naphthalate). Primary package decoration methods include labeling, sleeving, and direct printing. Labels are required to inform consumers, comply with legislation, and protect package integrity. Attention is given to achieve total adhesion, accurate application, and prevention of label damage and discoloration. The amount of information needed dictates the decoration method and legislation can determine the optimum text size. "From the perspective of legislation, it is absolutely essential that the graphics of the packaging should in no way be damaged or deformed by the application process, for example, distortion of graphics printed to a plastic sleeve during the shrinking operation." (Giles, 2000)

Non-shrinkable roll-fed labels are perhaps the most common label for use on beverage containers, and thus the most competitive alternative label style. After printing, the substrate is wrapped around a container and the ends glued or melted together to form a circular label.

Shrinkable roll-fed labeling applies a conventional wrap-around label capable of some shrinkage, typically less than 20%. A common use is to cover the curvatures at the top and bottom of aluminum cans with the label. The limited degree of shrink that roll-fed film offers prevents it being used for conical or tapered, highly shaped, or narrow containers. Polyethylene (PE) film's elasticity allows sleeves to be stretched over containers, where it reverts back to its original size to form a snug fit. Stretch sleeves too offer 360-degree decoration on plastic containers but are not suited for tapered, shaped, or narrow containers because the process requires uniformity of shape in decoration area. However, full body shrink sleeves (FBSS) are made from more elastic polymers and can be applied to these containers that require more label transformation.

The popularity of FBSS increased faster than sufficient understanding of the process and the potential of sleeves. The use of shrink sleeves, invented in Japan in the 1970's, is now a quickly growing and versatile container labeling method. (Giles, 2000) Polyvinyl chloride (PVC) was first used to produce shrink closures for wine bottle caps. It was then introduced with glass beverage bottles as a safety feature to retain broken glass. Eventually, the use of shrink sleeves entered European markets during the mid 1980's. The technology was not developed enough for individual containers and was used only for twin packs, promotional items, cosmetics, toiletries, and personal care products. Machine developments and substrate advances allowed the packaging community to realize greater potential for the technology. Today, FBSS are used as primary labels on curved container shapes, decorated multipacks, short run market tests and prototypes, and seasonal labels sold at point-of-purchase displays.

FBSS also allow labeled containers to be decorated all around, improving brand image and increasing shelf appeal and sales. At first, there was a tendency to follow labeling conventions, using shrink sleeves as a band, or as a combination front-back label with a neckband. Designers eventually realized their freedom to innovate container and label simultaneously, and use all sleeve space for labeling. The shrinking process means the original and final product dimensions must be considered. Where shrinking is greatest, graphics must be pre-distorted to compensate and it is best to design a label with this process and the final appearance in mind. The most effective and cost-efficient method for checking pre-distortion work is digital proofs. (Giles, 2000) For oval containers, the sleeve must be positioned correctly or label panels will face the wrong direction and distort incorrectly. Barcodes must be placed where they will distort the least, with the bar code lines parallel to the container base.

Substrate manufacturers have developed high shrink materials to accommodate the more highly contoured containers being produced, each with properties that are useful in different situations. There are typically three materials in use as FBSS. In order from lowest cost to highest, polyvinyl chloride (PVC), oriented polystyrene (OPS), and polyethylene terephthalate glycol modified (PETG). These materials are produced through either cast film extrusion or blown film extrusion. Essentially, plastic pellets are melted through friction and pushed through a die. The film is chilled as it continues to move along the production line (machine direction), through a series of rollers that maintain tension. Then comes the tentering process of production (Figure 3), where mechanical clips stretch the film wider (cross or transverse direction, perpendicular to machine direction). This elongates the film at a molecular level and any energy later imparted to the material through heat will cause the molecules to contract back to their original size. Thus, the shrinking property is built into the film. For all FBSS, shrink ratios are different in each direction, and usually substrates are oriented so they shrink most horizontally around the circumference (cross or transverse direction) with a small amount vertically (machine direction).

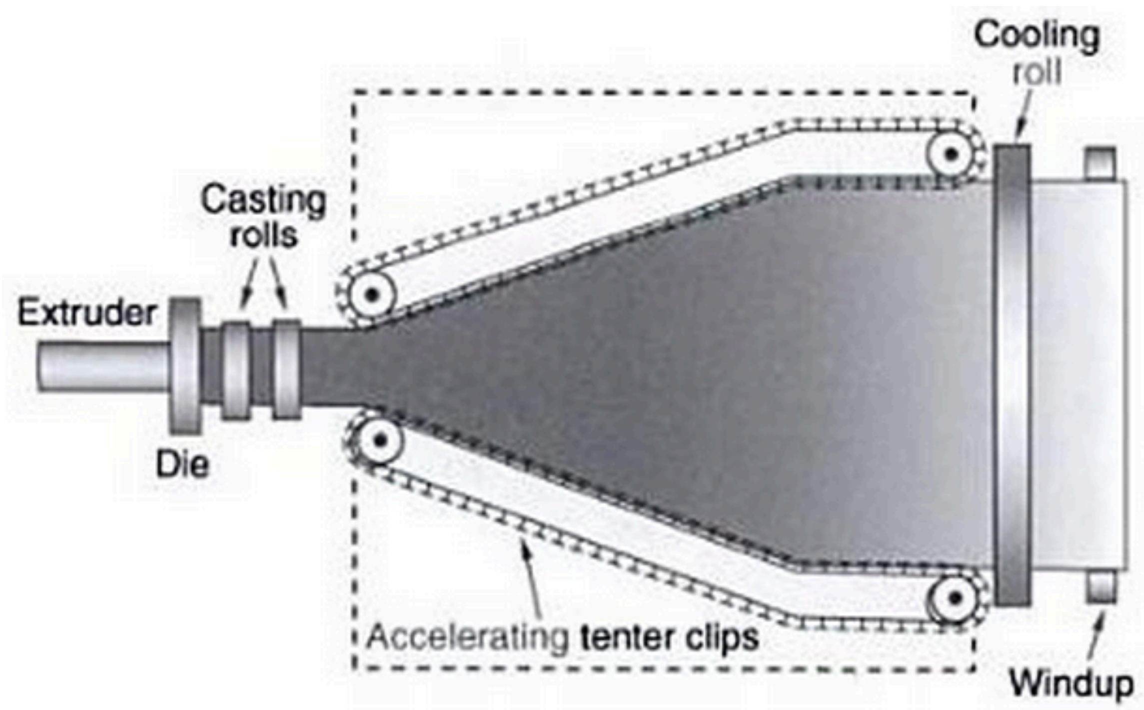


Figure 3. Tentering process of stretching film in the cross or transverse direction. (Robertson, 2006)

Hot air, steam, electrically heated elements, or infrared can all be used as heat sources to conform the sleeves to container profiles. (Giles, 2000) The applied heat should be directional, applied in greater amounts to high shrink areas, and kept to a minimum. Problems are most prevalent with the overall width of sleeves, particularly where container shape is less round which can cause snagging during sleeve application. During the shrink process, problems include cold air draughts causing uneven temperature in shrink tunnels, incorrect line speeds (or temperature) preventing appropriate shrinkage, container proximity preventing even air circulation, and containers cold from storage absorbing heat. Low temperature shrink film is useful for labeling plastic containers that otherwise might be in danger of deformation during the shrink process.

Because the seal must not break open during the shrinking process, the quality of the seaming operation is very important. The most common solvent is tetrahydrofuran (THF), which melts the edges of the film together. (Giles, 2000) Every sleeve is prepared by being laid flat, giving it two folds, and the seam is best placed halfway between the design center and one of these folds. Typically, images are printed on clear shrink sleeves in four or six colors with a white backing. The label is applied with the ink is on the inside or back side, so graphics are protected between the container and sleeve itself. This also allows the clear substrate on top to cause a gloss effect. Printing is usually done with the rotogravure process, though sometimes the flexographic process is used for simpler labels. Production speeds for fully automatic labeling machines range from 20 containers per minute (cpm) up to 1200 cpm, with many in-line machines maxing out at 350 cpm and rotary machines at 800 cpm. (Giles, 2000) Scuffing and scraping is less of a danger for plastic containers than glass during line production and transit, but slip lacquers or protective silicone coatings may be applied during printing, or by spraying the finished product right after the sleeve is applied.

Besides allowing full decoration on contoured containers, FBSS are gaining popularity due to compatibility with numerous features like metallic colors, matt and gloss finishes, mimicking acid etched glass with frosty finishes, scuff resistance, high opacity finishes for UV protection, pasteurization capability for pre-labeled bottles, UV ink that glows in the dark, translucent finishes, and thermochromatic and photochromatic inks. (Giles, 2000) Environmental concerns are impacting shrink sleeves and the packaging industry at large, and increasingly materials use sustainable sources and maximize recycling abilities. Lighter weight bottles allow more per shipment and a smaller consumption of transportation resources.

***FBSS Production Process***

The Fort Dearborn Company (FDC) is a leading producer of full body shrink sleeves (FBSS). According to Doug Becker, Technical Service Director of Roll Technologies at FDC, the selling points for FBSS are almost always the increased visual appeal, vast assortment of container shapes which may be utilized, and increased billboard space for the brand owner. FBSS typically are the highest cost decorating method compared to pressure sensitive or roll-fed labels. Most brand owners feel that the increased marketing potential overrides the increased production cost, and usually they experience a significant increase in sales when they switch to FBSS. (Becker, 2014)

A typical production run can be large or small depending on the method. FDC accepts minimum order sizes of 25,000 impressions, or labels, for the flexography print process. For the gravure print process they require a minimum order of 100,000 labels. Most customers who purchase FBSS from FDC shrink and apply them at an average rate between 40 to 160 cpm. Very few customers have equipment that can apply labels faster, and some customers will hand apply labels at relatively slow speeds of about 20cpm. The speed largely depends on container size and shape. (Becker, 2014)

Most FDC customers have either hot air or steam shrink tunnels. To achieve consistent quality shrinkage, the typical residence time of a bottle in the tunnel should be about 6-8 seconds. In a steam tunnel the average temperature is often 195°F. Hot air tunnels can be anywhere from 200° to 400°F depending on the length as shorter tunnels require higher temperatures to achieve adequate label shrinkage. Most FBSS by FDC are in the range of 40% to 65% shrink ratio. (Becker, 2014)

The first important factor in the FBSS process is shipping, storage, and handling of the sleeves before they go to production. Any exposure of the labels to temperatures higher than 80°F in this stage will cause sleeve dimensions to creep smaller. Even a 1mm reduction in the layflat dimension of a sleeve will negatively affect the performance on the sleeve application equipment. A common challenge in the actual application process is finding the proper coefficient of friction so that the sleeves feed well on the applicator, don’t jam, and go all the way down the container to the proper placement. The next area of difficulty is the tunnel, which can take a great deal of work to dial in. There are many parameters that require adjustment to achieve adequate, repeatable shrink. In hot air tunnels there are temperature, airflow, and conveyor speed settings in each zone to adjust. On a steam tunnel there are steam inflow and exhaust, nozzle positions, and conveyor speed. Every tunnel shrinks differently and it is very hard to fine tune two tunnels, so testing is done on the actual line on which the real product will run. (Becker, 2014)

Computers are only used in the graphics area of FBSS. FDC provides a sleeve with a basic grid pattern in the appropriate size for a customer to apply to their container. That grid-covered container will then be scanned into a software program at the graphics house to determine distortion maps, used to compensate by pre-distorting the final graphics. Not all graphics are distorted however. Some customers do not pre-distort but keep graphics out of high distortion areas. FDC recommendations for graphics are that circles are typically bad, boxes around text look poor, that shrink sleeve companies ought to counsel designers on what will work, and some regions on sleeves have such high distortion that any graphics are ill advised. It is difficult to produce/apply the label appropriately. A change in graphics occurs even in a successful production run. (Becker, 2014)

The process to set up FBSS begins with a customer sending sample bottles to FDC, who will then calculate the bottle dimensions. Of particular interest are the maximum and minimum circumferences. They select a material based on the shrinking requirements. They keep large stock of printed grid material, a continuous length to be cut to bottle height, which is used to measure the sleeve after it has been applied to the sample bottles. The specific square grid used can be 3mm, 5mm, or greater. FDC will measure graphics by hand but a competitor, Hamer LLC, prints hash marks used to calculate distortion by computers. The grid can then be imported in to graphics software, such as Esko’s Shrink Sleeve Toolkit, for pre-distortion and printing. The final preparation step is called the "disaster check," where labels are printed in one color, inspected, and then applied on containers until the shrinking conditions may be established. Afterward, the customer's FBSS production run may be manufactured. (Becker, 2014)

### *Human Visual System*

The importance of visual stimulation among human senses has been established (Bloch, Brunel, and Arnold 2003). There are two types of distortion measure used for picture quality prediction. The first type, the raw measures of error, give simple mathematical deviations between the original and a reproduced image, done mostly using computers. The second type of distortion measure is based on human perception. The human observer is the end user of most image information, and distortion is quantified as viewer dissatisfaction. (Karunasekera & Kingsbury, 1995) To understand how forms and distortion are received, knowledge of the Human Visual System (HVS) and memory are necessary. In the last twenty years, magnetic resonance imaging machines (MRIs) and other noninvasive functional imaging equipment have allowed us to map the human visual system during perception. (Grill-Spector, & Malach, 2004)

The HVS is very capable, able to process image content and identification in as little as 150 milliseconds with up to 94% accuracy. (Thorpe, Fize, & Marlot, 1996) The HVS can differentiate types of biological motion and even gender from extremely limited visual displays such as light-points that map joints in body structure. (Grill-Spector, & Malach, 2004) Processing in the HVS may be so effective due to a functional hierarchy, meaning there are many areas in the brain activated depending on the nature of a task. (Orban et al., 1996; Grill-Spector, & Malach, 2004) It is theorized that the brain utilizes cooperative processes for reading images in different conditions. One would be detection-based for high quality images containing near threshold distortions, where the HVS attempts to overlook the image itself, and any masking due to contrast and subtle lighting differences, to find distortion. The second would be an appearance-based strategy for low quality images, sifting through extreme distortion to pull out image content using indistinct edges and form analysis. (Hancock, Burton, & Bruce, 1996; Kanwisher et al, 1996; Larson & Chandler, 2010)

It has been argued that faces are a special category of objects with rapid recognition and processing efficiency. (Hansen & Hansen, 1988; Hancock, Burton, & Bruce, 1996; Puce et al., 1996; Thorpe, Fize, & Marlot, 1996) According to literature, faces evoke more brain cortex activity than non-word strings of letters, particularly in the creative right hemisphere. In a study by Puce et al (1996), texture was used as a non-object control, though it does share brain activation regions with processing colors, shapes, and scrambled faces.

Contrast is a primary sensory feature in HVS perception. When edge contrast increases it usually enhances perceptual quality, while blurred edges damage perceptual quality. (Lin et al, 2005) The HVS has frequency sensitive and frequency masking properties, meaning the visibility of artifacts is affected by local surroundings. In an example where the viewer is facing a crowd of people, facial forms become a background texture and the brain supposedly reverts to the second appearance-based strategy, processing faces as features in an array, and giving attention to features that stand apart. This filtering works best with "angry" faces, which act as the distorted feature of the background texture, especially the edge line of the furrowed brow. (Hansen & Hansen, 1988) This is perhaps because sensitivity increases with edge length. (Karunasekera & Kingsbury, 1995) In fact, downward-pointing angular, diagonal lines resembling a 'V' seem to be particularly representative of facial displays of threat. (Aronoff et al., 1988, 1992) Interestingly, distinct faces are not the same as memorable faces, in part because these two characteristics are disassociated in their processing. (Hancock, Burton, & Bruce, 1996; Isola et al., 2011) If a package is distinct it may have an initial advantage, though a memorable package may be better able to make long-term connections with consumers.

***Markets, Memory, & Illusion***

Before studies by Isola et al. (2011), there were no databases that categorized images by degree of memorability, which would be useful in many applications. They used Amazon's Mechanical Turk (MTurk) website for visual memory studies presented as games and found that, consistent among different viewers, what makes an image memorable is a stable intrinsic property tied to what makes it meaningful, the objects it shows and the scene category. This is true even without familiar artifacts like family relatives and iconic monuments. (Isola et al., 2011) Perceptual features like mean hue, saturation, and value do not seem to be retained in long-term memory and are at best weakly predictive of memorability.

Image structure is gathered by the HVS in "coarse-to-fine" detail sweeping, where broad feature distinction and known-feature correlation combined can categorically identify an object, after which narrow features advance specification. (Clarke et al., 2012) The stronger the feature correlation, the faster recognition occurs. As this transpires, processing advances to the frontal lobe, responsible for coordinating information including that of memory. If an object is distinct, it may be because it has less meaning and does not correlate as strongly to anything in memory. This could mean an object with many complex features requires more correlation, is harder to perceive, and is less likely to be meaningful or memorable. For example, when a package is cylindrical with a horizontal to vertical ratio between [3:5] and [2:5], features of logo typography can significantly affect findability and legibility. In this range, sans serif fonts with equal stroke thicknesses are more easily found and perceived, more commonplace and memorable. (Chou & Wang, 2011)

"Word recognition requires identification of letters while also the spatial arrangement of letters." (Yu, et al. 2014) Mental word processing moves from raw sensory information to feature segmentation to recognition of letters and position. "Crowding" refers to the hindering of visual comprehension due to interference from neighboring objects. Common with peripheral vision, crowding usually results in mixing up the position of letters. Reading speed, visual-span size, and the effects of crowding depend on the properties of text and the distance from the center of vision. In a study by Yu et al (2014) reading speed for marquee text, upright letters arranged in a vertical column, was 42% of horizontal reading speed, and the reading speeds for text rotated 90 degrees clockwise or counter-clockwise was 55% of the horizontal speed. They concluded that slower reading is due to reduced visual-span size and greater vertical crowding.

On the other hand there are levels of memory that can explain brain speed and efficiency; the weaker recognition and the stronger recollection memory. (Gardiner & Java, 1990) Recognition is weaker because it requires prompting. Memorable images generally include people (faces), interiors, foregrounds, and human-scale objects, all things meaningful to the every day experience. Images of landscapes and peaceful, aesthetically pleasing or unusual scenes are not memorable. (Isola et al., 2011) The application of typicality (recognizable or memorable images and objects fit expected categories) could impact how consumers receive and respond to new graphic design trends, to shrink sleeve labels on certain products, or to certain bottle shapes. (Westerman et al, 2013) Chou & Wang (2012) recommend that future studies take into account degree of memory, cognition, and attention rather than focusing only on design elements.

Findings about design are largely generalizable across different packaging stimuli, including logos, containers, and marketing elements. Orth & Malkewitz (2008) were the first to assemble systematic design elements and trace how prototypical holistic package designs are rooted in those elements. They believed design elements influence brand impressions, processing fluency, brand recognition and recall, and described five packaging “personalities” that plausibly exist for a wide range of products: massive, contrasting, natural, delicate, and nondescript package designs. (Orth & Malkewitz, 2008)

Products packaged in plastic bottles are now very common among consumer goods. (Chou & Wang, 2012) The proportion of length to width of packaging, or the contours of plastic bottles, will affect preference for a product and thus market sales. (Raghubir & Greenleaf, 2006) Naturally, there has been a recent proliferation of subtly anthropomorphized beverage container shapes, a "skinny" physique, broad angular "shoulders," or a curved "waist," that may differentiate the products, reduce volume and costs, or make the bottles easier to hold and use. Anthropomorphized package shapes are associated with specific traits that consumers incorporate into their identities, especially when traits can be instrumental in goal achievement. (Schneider, 2014)

Men are more image-driven in their preferences for specific designs and more concerned with the overall structure of a product, whereas women pay more attention to details, textures and natural-looking forms. Both genders value design simplicity, although men view it in image-related terms, while women prefer it in practical terms. (Xue & Yen, 2007) A study comparing 3 bottle forms by Grahl, Greiner, & Walla (2012) "provides evidence for gender-specific unbiased emotion differences related to different bottle shapes." In their study, both genders had a strong preference for bottle 3 (88%), a slim, tall and pyramid-shaped bottle. Bottle 2 was rated least preferred (81.25%), especially by men. They concluded that bottles 1 and 3 did not resemble any bottle commonly used for the storage of liquids, but that bottle 2 resembled one containing strong alcohol like whiskey, with which men may have had stronger negative associations and experiences.

Psychological factors can determine reaction to advertising images. In advertising, people with high self-esteem respond to attractive, confident models while those with low self-esteem respond better to images of products themselves, which aligns with motives of verifying or enhancing self-esteem. (Aydinoglu & Cian, 2014) Studies on body-image perception and attitude have concluded that people with eating disorders overestimate body size more than control subjects, with both body-site and whole-body images. (Gardner, 1996) Face recognition has also been demonstrated to be unaffected or even enhanced by some types of systematic distortion, such as with caricatures. As early as 7 months of age, infants seem to develop adult-like face recognition, which includes preference for their own mother's face even when it has been vertically stretched, indicating horizontal dimensions may be more related to face recognition. (Yamashita, 2014) Relating more broadly to psyche and perception, researchers Rachman and Cuk (1992) reported that people with acrophobia exaggerate estimates of height, and that successful therapy for their phobia reduces overestimation. Consumer psyche may cause them to react quite strongly to packaging, which they find fearful or relatable, or if it simply includes a face.

Several context effects exist that may play a role in consumer behavior. The similarity effect states that when three products are present, if two are similar they will compete for consumer interest while the third will remain just as likely to be selected. With the attraction effect, if a third product introduced is just slightly less favorable than the first product, the first will seem far better than either the newly introduced product or the second by comparison. The compromise effect says that when a third product is introduced that makes the second appear to be an average of the first and third, consumers have an increased probability of choosing the middle product as a compromise. The reference point effect is the last context effect and is not yet as well understood. It says that when a consumer has experience with one product and is deciding between two new products, they will tend to select the most similar new product in terms of tradeoffs, opting for one with small advantages and no disadvantages over those with large advantages and large disadvantages. (Busemeyer et al., 2007)

Pricing research by Orth et al. (2010) has found that consumers form pricing expectations based on visual cues of brands. "Consumer goods firms spend billions of dollars each year to appeal to buyers by inventing, engineering, and manufacturing package designs" to set their products apart. "Research indicates two routes of consumer package evaluation, a central or cognitive route via quality inferences and a peripheral or affective route via attractiveness." Designs that aim for such qualities as "natural, harmonious, or elaborate" may influence consumer price expectations directly or indirectly. (Orth et al., 2010) Package design seems to have the ability to influence consumer price expectation through consumer judgments of quality and attractiveness. To some extent, aesthetically appealing packages can represent a source of value to aesthetically conscious consumers and may even lead them to accept higher prices.

Effective logos are simple, memorable, timeless, versatile, and appropriate. (Cass, 2009) Defining "good design," Hertenstein et al. (2013) used questionnaires to glean the perspectives of both corporate and consulting industrial design managers. They found that design is evaluated in a similar fashion to the decision-making process, utilizing categories and rules of comparison. Primary design analysis relates to customer experience, while others relate to business results that can, at times, require trade-offs between competing goals and interests like quality versus expenses. Company focus should be balanced and is in fact synergistic with customer experience.

Cultural differences influence how people perceive, respond to and prefer design qualities, particularly with brand impressions. (Aaker, Benet-Martinez, & Garolera 2001; Hekkert & Leder, 2008) Designs of brand logos tend to be angular rather than rounded, and particularly so in Western, individualistic cultures (Westerman et al., 2012) Ideally, a design should clearly convey the same intended meaning among all peoples, and not be easily misinterpreted. To market across cultures, it is often a matter of figuring out how to best say a simple message (Keller 2003; Underwood 2003). Logos are the repositories for brand associations, their design and selection costly in terms of time and money, and they are the most common marketing element to be used unaltered when going abroad. Alternatively, Asian brands have often struggled to develop quality images. (Djurovic, 2009) Coca–Cola has led the way in packaging localization. For example, overseas the Chinese typeface has been integrated into the brand identity and used on Coke's visual communications and ads there. In studies by Henderson et al. (2003) there were sufficient similarities between China and Singapore to suggest that managers can use a single visual strategy to achieve strong positive responses across countries, and that Asian brand symbols could be transferable to the United States.

In previous studies, quantifying subjective human emotion or preferences has been done through questionnaires or keyword associations. But while still influencing behavior, basic components of emotion may not be conscious. Explicitly stated answers to questions about subjective preference may not result in reliable information regarding non-conscious emotional aspects. (Walla et al., 2002; Walla et al., 2003; Winkielman & Berridge, 2004; Walla, 2008) Henderson et al. (2003) developed a hypothesis that "responses to design are a function of non-consciously acquired internal processing algorithms" formed by exposure to a series of objects. These hypothetical rule systems that govern preferences fall into three categories: perception, motivation, and cognition. Holbrook & Schindler (1994) agree that perception of design aesthetics sways preferences. Gray et al (2002) and Ochsner & Phelps (2007) support that emotion-related information can motivate even decisions seemingly made from a rational perspective.

Gestalt psychology, recent design theory, and empirical evidence suggest consumers perceive "constitutive" elements and organize them into more complex whole components (Veryzer, 1999). This hierarchical processing draws from the HVS, studies on non-conscious preference formation (Kunst-Wilson and Zajonc, 1980; Lewicki, 1986), and findings that abstract, multidimensional design qualities are determined by basic, measurable design elements like unity, symmetry, and simplicity. (Geistfeld, Sproles, and Badenhop, 1977) Consumers move from initial design perception to interpretation, trying to understand and place a stimulus within an existing category (Loken & Ward, 1990). Visual attention filters sensory information during this processing. A shift in visual attention, known as a saccadic eye movement, alters spatial and temporal perception by piecing together separate focal points into a continuous, comprehensive mental image. This means the HVS can miss things between saccadic eye movements, or perceive things without conscious visual awareness of them. Thus, an observer may not consciously report anything below this explicit “phenomenal awareness” level, though subliminal stimuli have an effect. (Au et al., 2013)

Attention is not the only thing that changes what is perceived. In a study by Wesp et al., (2000) people estimating the size of familiar objects made significantly larger errors than those estimating the size of undefined objects. In a second study, size estimation errors from memory were larger than when objects were directly viewed. Experience with objects appears to decrease accuracy of size estimates and influence accuracy of the perceptual representation of physical attributes. The findings of these two studies also support that what an object is may influence perception of its size. Size may be an intrinsic element of an object’s mental representation based on one's first experience with that object, and that size is encoded "relative to the perceiver." Errors of size judgment from memory may stem from the original assessment of size, from memory of size, and from new experiences with the object. Wesp et al. believe orientation of the object may influence size judgments. Figure 4 includes the Ponzo Illusion, where the HVS estimates the size of an object based on its background.

It is postulated that abstract characteristics are object-centered and surface characteristics are viewer-centered, or that abstract object-level representations are distinct from surface view-specific representations of visual stimuli. It is possible that size-specific information is retained in a different format in long-term memory. The HVS might know an object has a specific size but not use this knowledge to form the object's visual image in long-term memory, a size not linked to the observers' point of view. Iachini & Giusberti (1996) assert that only certain levels of representation in the HVS preserve size-specific information.

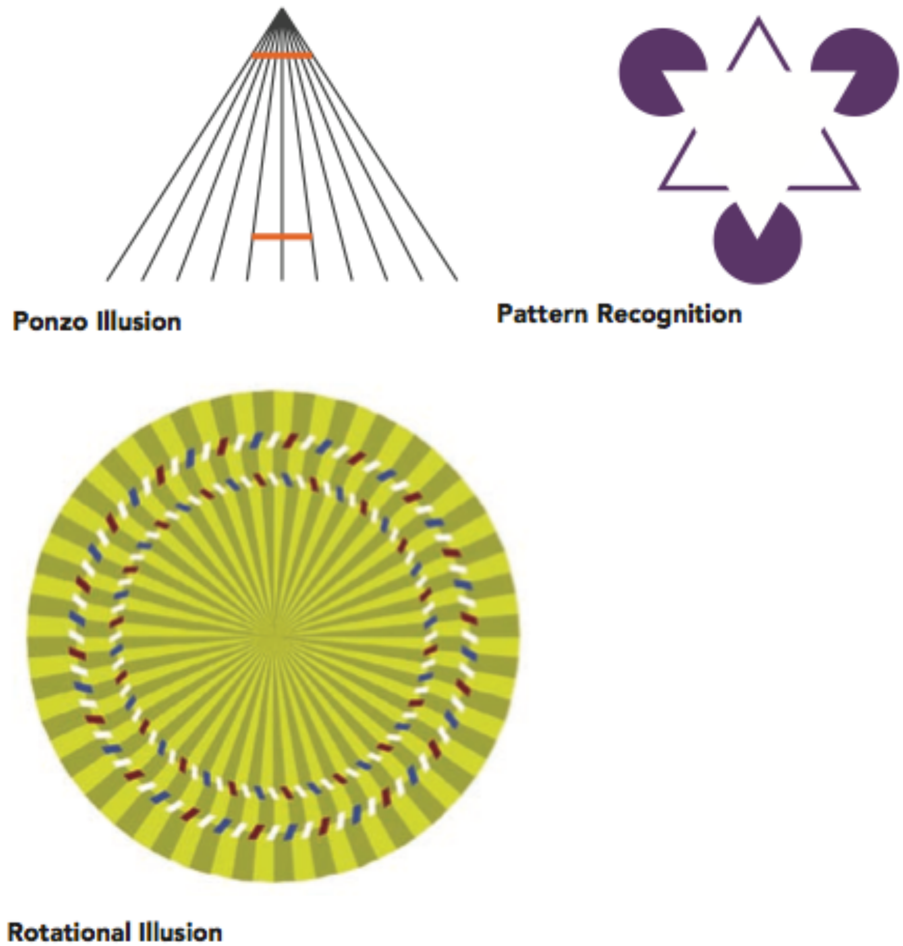


Figure 4. Basic visual illusions that have implications on perception. (King, 2014)

### *Computer Image Quality Assessment*

Distortion can have a different meaning among various fields of study. Work has been done to allow computers the same image processing capabilities as the HVS, evaluating images with the same criteria as people would. (Wang et al, 2004). "Perceptual visual quality metrics (PVQM) aim to quantify the quality of visual information, including still pictures and video. It can be extended to 3D models, computer graphics, animation, and 3D and multi-view visual data. This is an interdisciplinary field involving vision science, color science, signal processing, physiology, psychology and computer engineering." (Lin & Kuo, 2011) PVQMs provide objective metrics during all design phases and reduce human subject evaluation. Research has gone into finding various approaches of objective image quality assessment and writing computer algorithms to perform said operations. The six most used image metrics include SSIM (structural similarity index), VSNR (Visual Signal-to-Noise Ratio), IFC (information fidelity criterion), VIF (visual image fidelity), MSVD (method of surface virtual dislocations) and PSNR (peak signal-to-noise ratio).

"PVQMs can be classified with regard to reference requirements: double-ended and single-ended. Double-ended metrics require both the reference (original) signal and the test (processed) signal, and can be further divided into two subclasses: reduced-reference (RR) metrics that need only part of the reference signal and full-reference (FR) ones that need the complete reference signal. Single-ended metrics use only the processed signal, and are therefore also called no-reference (NR) ones." (Lin & Kuo, 2011) The HVS is still much better than computer algorithms at comparison with reduced and no-reference methods. (Pappas, Safranek, & Chen, 2000; Sheikh & Bovik, 2006)

Not every change in an image leads to noticeable distortion, because not every area of an image receives the same attention level. Due to masking effects, changes are often not fully perceived. (Lin & Kuo, 2011) The just-noticeable-distortion (JND) threshold refers to a point below which a change cannot be detected by a majority of people, typically 75%. If a change is below the JND value, it can be ignored in visual quality evaluations. An original or ideal image is used as benchmark, and metrics are adjusted to compensate for error. This is done without surpassing the JND or masking threshold, beyond which the error would be noticeable. (Lin, & Jay Kuo, 2011; Wang, & Bovik, 2002)

External factors and masking effects are accounted for in PVQMs by viewing conditions, such as ambient illumination, display resolution and viewing distance. The effect of viewing distance is actually related to display resolution, though there has been limited research on the influence of the viewing distance, and the issues related to ambient illumination are largely uninvestigated. (Lin & Kuo, 2011) Viewing distance is a factor of visibility and the minimum viewing distance is where distortion is most detectable. (Pappas, Safranek, & Chen, 2000) Effort is being made to develop universal image quality metrics, meaning the same standards could be used for even unrelated images. Temporal masking of distortion is considered when evaluating video, so when testing consecutive static images time and visual separation should be allotted for best assessment. Pertinent to the present author's research, this translates to self-pacing surveys and vacant interval slides to prevent carry over affects between stimuli presentations. In spite of progress in physiology, psychology, HVS research, and computer science, PVQMs matching human perception is not yet fully attainable due to the complex, multi-disciplinary nature of the problem. (Lin & Kuo, 2011)

***Distortion Correction***

For specific image elements and applications, distortion of form in moderation is desirable, if not necessary. Depending on the purpose, greater degrees are acceptable and required, though not explicitly noticed. (Agrawala, Zorin, & Munzner, 2000) Thus, distortion is not to be avoided but rather understood and exacted to correct or create visual effects. The mechanics determining automation of this control, of interest to various industries, are based on perceptually preferable tradeoffs. These tradeoffs weigh structural features, deemed the essentials of representation and perception, and determine the minimal procedures to prevent and counteract distortion. (Wang, Bovik, Sheikh, & Simoncelli, 2004; Zorin, & Barr, 1995) One measure of structural features is how well the zero-curvature condition is satisfied, which looks at the maximum curvature of a line segment that should be straight. A function calculates this based on a point along the line, the line direction, and the image size. When distortion error is evenly distributed in all directions, a rare condition except with wide angle and fish eye lenses, a central symmetry function may be applied. With computer graphics and animations, it is generally used as a simplified solution in place of a non-symmetric error correction. (Zorin & Barr, 1995) Texture has a dependent relationship with surface curvature that requires a more involved distortion correction with photo editing software. Essentially, this process repositions curves, generates texture from source, and synthesizes the two with appropriate texture orientation and scale. (Fang & Hart, 2007) Using a contrast sensitivity function (CSF), edge contrast increases usually enhance visual quality because pixel errors are much easier to detect along edges than in textured regions of an image. (Lin et al, 2005)

Often, minimal error correction is as preferable as minimal distortion, so preservation functions transform most where an image has the least detail, coordinating local changes in a comprehensive procedure. (Karni, Freedman, & Gotsman, 2009) Two applications from this are in aspect ratio transformation and image-sensitive object deformation. More methodical Moving Least Square (MLS) functions apply a grid to an image, deform at vertices, and readjust within quadrants, reducing affected pixels. (Schaefer, McPhail, & Warren, 2006) This can help retain smoothness in an image. If distortion is predicted to occur later, as it is in the shrink sleeving process, proactive image manipulation using software based on these techniques can be a compensative solution.

***Sensory Evaluation – Threshold Testing***

Psychophysics utilizes thresholds, observer analysis, and signal detection theory (SDT) in "the scientific study of the relation between stimulus and sensation," or the study of perceptual systems. (Gescheider, 1997) Thresholds, also called points of subjective equality (PSE), are the measured limits of sensitivity to small changes in stimuli. (ASTM E1808) They are used with many goods and applications to determine the point at which known contents or qualities begin to reduce acceptability. They are frequently difficult to determine and reproduce, but may still be the best approach available.

There are four types of thresholds. The absolute or detection threshold is the lowest stimulus capable of producing a sensation. This includes “the dimmest light, the softest sound, the lightest weight, the weakest taste.” (Meilgaard, Civille, & Carr. 1999). Next is the recognition threshold, the minimal amount of something that may be both recognized and identified. Thirdly is the just-noticeable-distortion (JND) threshold, the increment of stimulus change required to produce a perceptual change. The first JND along a spectrum, the first perception of change above no stimulus, is also the absolute threshold. (ASTM E1808) Finally, the terminal threshold is the point at which any increase in stimulus magnitude will not increase perceived intensity of the quality of that stimulus, and pain is often observed beyond this point for some senses. The aim of this thesis was to identify the first JND or absolute threshold for graphic distortion in FBSS. This would mean consumers know graphics are distorted but cannot pinpoint where or by how much.

A threshold is typically thought of as an absolute, above and below which a stimulus either is or is not detectable. In reality, a person’s sensitivity fluctuates with environmental and physiological differences. A threshold is not a constant for a given stimuli, but a constantly changing point on the sensory continuum from non-perceptible to easily perceptible. It can change with mood and biorhythm, hunger, and other variables. (Meilgaard, Civille, & Carr, 1999) Thresholds are not constant between stimuli either. Even stimuli with identical recognition thresholds can increase very differently in perceptual intensity with increased concentration, measured by successive JND thresholds. Thus, the use of thresholds as a measuring stick must be done with caution and understanding. (Meilgaard, Civille, & Carr, 1999) Two types of responses can be used in threshold testing: yes-no (or pass-fail) and forced-choice. (ASTM E1808) For pass-fail experiments, the threshold is computed as the stimulus value that is detected 50% of the time. When 50% correct response rate represents chance behavior, such as with the 2 alternative forced choice (2-AFC) test seen in Figure 5, the point between chance response rate, 50%, and perfect response rate, 100%, is usually taken as sensory threshold, 75%. (Gescheider, 1997; ASTM E1808) Individual thresholds may be used for calculation of a group or panel threshold. The size and composition of the panel (usually 5 to 15 members, preferably more) is determined according to the purpose for which the threshold is required and the limitations of the testing situation. (ASTM E1432-04) Pooling of the data sets from panel members to produce a single step calculation of the panel threshold is not permitted. The panel threshold must be calculated after individual thresholds to account for variance between panelists. A panel sensory threshold is most often normally distributed, but skewed right due to most groups having individuals with low sensitivity or high indifference to the stimulus. The geometric mean is the measure of central tendency best suited for identifying the group threshold as it is least skewed by the highest values from personal thresholds. (ASTM E1432-04)

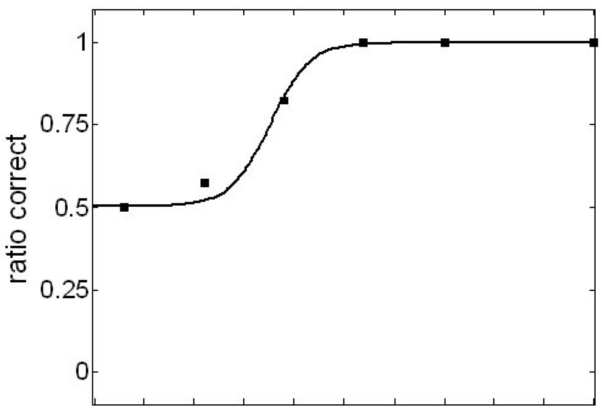


Figure 5. Range of probability with 2-AFC Method. In forced-choice experiments, the point between chance response rate, 50%, and perfect response rate, 100%, is usually taken as sensory threshold, 75%. (Gescheider, 1997; ASTM E1808)

Psychometric functions can be derived for either a single observer through multiple trials or a population of observers with one or more trials per observer. Rather than responding to stimuli, the error of habituation (responding from habit) and the error of anticipation (responding from prediction) both interfere with psychometric research. These must be counteracted by careful sample preparation and presentation. Classical psychophysical designs to measure thresholds are the Method of Adjustments or Average Error (MAE), the Method of Limits (ML), and the Method of Constant Stimuli (MCS). The MAE has test subjects adjust a comparison stimulus to match a standard stimulus, with the average setting for the comparison stimulus being the threshold and the difference between this threshold and the standard stimulus being the average error. With the ML, a stimulus is adjusted up and down in a stair-step fashion until the test subject reports that they detect it or no longer detect it, respectively. The MCS, alternatively the

Method of Right and Wrong Cases or Frequency Method, was first described by Friedrich Hegelmaier in an 1852 paper. (Gescheider, 1997) It is similar in many ways to the Method of Limits, but stimuli are presented in a random order. This prevents the subject from being able to predict the level of the next stimulus, thus reducing errors of habituation and anticipation. (Gescheider, 1997) Probabilistic modeling, also known as Thurstonian modeling, considers the efficiency of the central processing that takes place in the brain during comparative judgments. It was a precursor to the development of SDT and is based on the ideas of variation of product perception and what can be called the decision rule or cognitive strategy. (O'Mahony & Rousseau, 2002) Recently, psychophysicists have tried applying SDT, measuring thresholds by the psychological perception rather than the measure of the stimuli intensity.

Both the American Society for Testing and Materials (ASTM) and International Standards Organization (ISO) first recommend the 3-alternative forced choice (3-AFC) method of sample presentation in which three samples are presented, two controls and one containing the stimuli under test. The ASTM's rapid method (E679) aims at determining a practical "best estimate" value close to the threshold, based on a minimum of testing effort, allowing larger panels. The ASTM's intermediate method (E1432) proceeds to determine individual thresholds by the concentration detected 50% of the time, and the group threshold by histogram of individual thresholds. ISO Standard 13301 is essentially a combination of the rapid and intermediate ASTM procedures. (Meilgaard, Civille, & Carr, 1999).

Scales of stimuli intensity are typically segmented into 4-10 levels, with each step differing from the previous step by a factor between 2 and 4 (ASTM E1432-04). Types of scales include Nominal (values are different by name), Ordinal (values are sorted by rank), Interval (values differ in degree along a scale), and Ratio (scale possesses a meaningful, non-arbitrary zero value). Most measurement in physical sciences and engineering is done on ratio scales (energy, light, heat, electricity, sound). Very informally, many ratio scales can be described as specifying "how much" (magnitude) or "how many" (count). (Sage & Goldstein, 2010)

Ernst Heinrich Weber, Gustav Fechner (of the Golden Ratio), and Stanley Smith Stevens all contributed to how psychometric discrimination tasks are conducted. Weber's Law states that the JND threshold between two stimuli is proportional to the magnitude of the stimuli. In other words, as a stimulus increases, greater change will be required in order for change to be perceived. Weber's law generally holds true for most sensory stimuli. (Sage & Goldstein, 2010) "Fechner's Law was the first statement of the now-accepted fact that the relationship between sensory experience and stimulus intensity is nonlinear." Stimulus intensity has a proportional, logarithmic relationship with subjective perception, meaning that if a stimulus has a geometric progression (multiplied), and perception has an arithmetic progression (additive), we can use scales to refer between the two. (Sage & Goldstein, 2010) Steven's Law, also called the Psychophysical Power Law, states that any stimulus not following Fechner's Law follows a power function with a different exponent, allowing for dramatically different graphs of perception. These may be seen in Figure 6. (Sage & Goldstein, 2010)

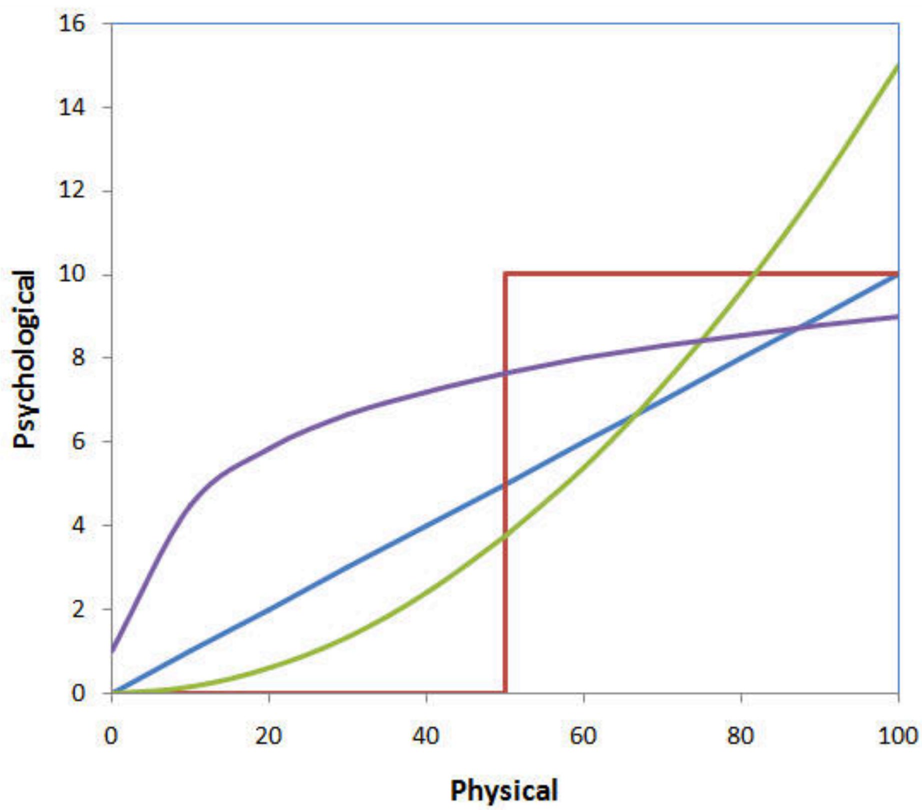


Figure 6: Functions of psychophysical stimuli.

Red: Common misconceived notion of a threshold.

Blue: Perceptual increase of stimuli according to Fechner's Law.

Green: Actual logarithmic increase in stimuli according to Fechner's Law.

Purple: A stimulus not following Fechner's Law follows a power function with a different exponent, according to Steven's Law. (Sage & Goldstein, 2010)

***Against the "No Preference" Option and 2-AC Testing***

The 2 alternative forced-choice (2-AFC) test, sometimes called the paired-comparison difference test, is generally used in discrimination-based psychophysics experiments, and requires test instructions to specify the nature of the stimuli attributes being discriminated. (O’Mahony & Rousseau, 2002) During a single 2-AFC trial, two stimuli are presented. In an absolute threshold experiment, one of the stimuli is a stimulus at 0 intensity, or no stimulus, and the other is the stimulus at some scaled intensity. The two stimuli can be presented in different locations, simultaneously, or in the same location, consecutively. If presenting the stimuli simultaneously, it is decided randomly which stimulus is in each location. If presenting consecutively, it is again decided which stimulus is presented first. (Green & Swets, 1966)

In classical methods, participants are typically asked about detection while being presented with one stimulus at a time, the intensity of which increases and decreases in a stair-step fashion around a predicted threshold. The task in a 3 or 2-AFC experiment is instead based on detection from among alternatives, so context has a larger role. If stimuli are presented simultaneously, the task is to indicate which is the target stimulus possessing some degree of intensity. If the stimuli are presented consecutively, the task is to indicate if the target stimulus is presented first or second. In both cases, if unsure when or where the target stimulus is, the participant must guess. The threshold is the stimulus intensity that can be correctly identified 75% of the time. (Gescheider, 1997; ASTM E1808) This is considered a pure measure of participant sensitivity without any input from a cognitive bias.

With non-forced choice, observers use their sensitivity, a criterion element, to make a decision. A criterion is an uncontrolled factor in the experiment, differing between and within observers as a form of response bias, and is a central problem in discrimination testing. (O’Mahony & Rousseau, 2002) Experience or training acts to increase sensitivity and decrease a subject's criterion, and thresholds obtained with such panels are usually lower than the threshold most consumers can detect. (Prescott et al, 2005) Also, participants may actively seek to find the lowest threshold in order to appease experimenters and distinguish themselves, often resulting in more false detections. During forced-choice experiments, observers often claim they do not perceive anything but still they can produce results above chance because they are not aware of all sensory information available. "The common reason given for this is that an unsure and less confident judge, who can distinguish between two products, may be reluctant to declare that he can do so." (O'Mahony & Rousseau, 2002) Forced-choice judgments do not require conscious experience, or decision about detection, and benefit from subliminal perception so results are unaffected by criterion differences, eliminating response bias. (Green & Swets, 1966; Braun et al., 2004; Alfaro-Rodriguez et al., 2007) Discrimination tests should be matched in sensitivity to normal conditions of consumption or use, meaning the most sensitive test is not necessarily best. (O'Mahony & Rousseau, 2002)

Furthermore, much has been debated on the necessity or even the suitability of a 2-AC (2 alternative choice) method as it is not as well developed or well understood as the ordinary forced-choice discrimination and preference method. (Alfaro-Rodriguez et al., 2007; Christensen, Lee, & Brokhoff, 2012) It has not been used as extensively in research, though it often seems an appealing substitute. (Braun et al., 2004; Alfaro-Rodriguez et al., 2007) The 2-AC method includes a neutral "No Difference" or "No Preference" option in an effort to know if consumers really are not able to discriminate or do not have a preference, information that may be imbedded in forced-choice data. Without the “No Preference” option, analysis is a matter of simple binomial statistics. By including a “No Preference” option, analysis of the resulting trinomial data becomes more complex. (Marchisano et al., 2003)

Some authors have argued for the omission of the “No Preference” option, and while various ways to square analysis of “No Preference” data with binomial statistics have been devised, these approaches are not necessarily endorsed. (Alfaro-Rodriguez et al., 2007) They include discarding the “No Preference” responses, splitting them equally between the two preference options, splitting them proportionately between the preference options according to the respective preference frequencies, and assigning the “No Preference” responses to one of the two preference options by coin toss. These options discard or alter data to fit multinomial responses to a binomial statistical analysis, but only if there were a negligible amount of “No Preference” responses, which is not always the case, would these procedures be harmless to the final experimental conclusions. (Alfaro-Rodriguez et al., 2007).

Use of a placebo comparing identical products opens up another approach to statistical analysis, though a placebo seems only to make sense when including a “No Difference” or “No Preference” option. (Alfaro-Rodriguez et al., 2007) Instead of comparing the test frequencies with the placebo frequencies, it would be simpler to examine only the test frequencies for consumers who reported “No Difference” or “No Preference” in the placebo comparisons. In this way, only data for unbiased consumers are considered, but the sample size is likely to be drastically reduced, again because more false detections result from participants attempting to appease experimenters and distinguish themselves. Even when presented with a placebo, 70-80% of consumers regularly express a preference and only 20-30% will report “No Preference.” (Marchisano et al., 2003)

# CHAPTER THREE

# PILOT STUDY

Oyama et al. (2008) reported relatively small cultural differences in the associations with abstract two-dimensional forms, making them useful for general studies of the HVS. Bell et al. (2014) used the radial frequency patterns (polygons) seen in Figure 7 as stimuli to test brain mechanisms for lower form processing, for features like curvature, size, orientation, polarity, and luminance contrast. Polarity means using dark elements on a light background or vice versa, while luminance contrast means the ratio difference between these hues. They found evidence for two methods of processing in the HVS: feature-selective and feature-agnostic global shape mechanisms, meaning the brain can interpret features together or independently. In the same spirit as these studies, the present author developed and conducted a pilot study to define an absolute threshold, the point where under optimal conditions subjects can perceive the least amount of graphic distortion in basic polygons.

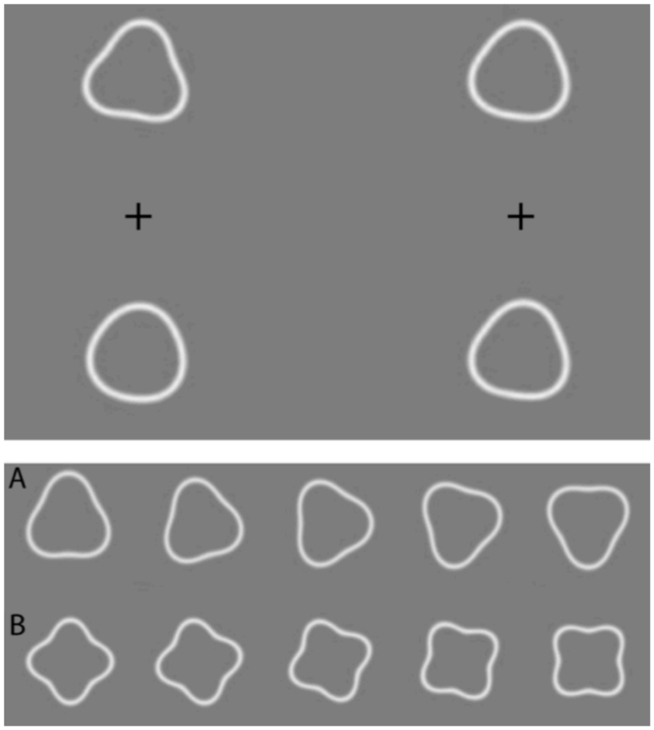


Figure 7. Radial frequency patterns, or polygons, used in case study. (Bell et al., 2014)

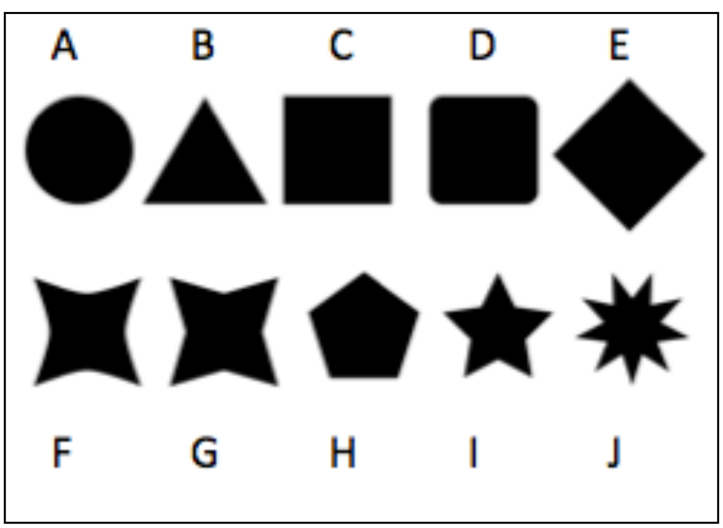


Figure 8. Polygons used in Pilot Study.

### ***Materials & Methods***

For this pilot study, the hypothesis was that "there is an absolute threshold at which a polygon is perceivably distorted and distinguishable from an original." The null hypothesis was then that "there is not an absolute threshold at which a polygon is perceivably distorted and distinguishable from an original." In designing the survey structure and length, the aim was to eliminate any learning curve and fatigue over the testing period. Another objective was to maximize the probability and ease of perceiving distortion, without encouraging false positives, in order to establish if there is a baseline threshold at which the HVS can detect graphic distortion. If found, then under other circumstances the threshold would be elevated with greater distortion required in order to be detected. The polygons seen in Figure 8 were used, as they are very simple and controllable forms. More complex shapes may carry connotations and various elements in larger compositions can produce distracting, masking effects. Symmetry is a form of balance and when a composition is built equally around a central point, it demonstrates radial symmetry. (Lidwell, Holden, & Butler, 2010; Bell et al., 2014) This simplicity and equality of form permits no bias to be read into the perception of an image. If the dimensions of a polygon do not appear uniform, it cannot be attributed to any inherent asymmetry and distortion must have been applied.

Proportion is the relative distribution of form to a whole. As with symmetry, proportion can play a role in how form is perceived. Next to other forms, features may appear to stretch or shrink, advance or recede visually. This informed the decision to use a 3-AFC test, with two identical and one similar polygon lined up horizontally, with no other images or information nearby to detract from them. Having more than three polygons within a comparison is perceived as too large of a mental task. (Lidwell, Holden, & Butler, 2010) Participants were asked to identify the polygon that was unlike the others, allowing them simultaneously to know the polygon's original appearance and attempt to identify an outlier. Similarity and proximity help our minds understand relationships between and within groups well. (Lidwell, Holden, & Butler, 2010; Orban et al., 1996) Presenting a solitary image does not always guarantee that a participant knows the original form, especially with less familiar polygons. (Gardiner & Java, 1990) One might think horizontal alignment would aid in identification of an odd form, but if placement were random, then proximity would be a compounding, uncontrolled variable. With horizontal alignment, saccade-type eye movement naturally follows a left-right reading motion and also follows the placement of like-products on a shelf in a shopping environment.

Color of polygons was consistently held to standard black, which eliminated texture and color bias, confirmed adequate polarity, and maximized luminance contrast between figure and ground, that is, the polygons and white background. (Karunasekera & Kingsbury, 1995) Contrast is the keystone structural feature, which aids the HVS in detecting form. (Lin, Dong, & Xue, 2005) A regular rhythm was maintained by giving the negative space the same lateral dimensions as the polygons. This eliminated inconsistencies in spacing, which would give distortion away through the visual closure effect, where the brain studies the interplay of edges and negative space. Because of the visual closure effect, it was important to include several edge conditions, corners and curves, which may be factors in distinguishing distortion. Corners may act as anchors for areas of interest (AOIs) and contribute to fixations. Furthermore, orientation of polygons was maintained, as it may impact perception of distortion by slowing identification of structural features. (Orban et al., 1996; Bell et al., 2014)

The pilot survey presented the polygons with varying degrees of distortion, from 0-10% less than original width. In all, there were ten polygons, beginning with a circle and increasing the number of features, points and sides. All polygons fit within a two-by-two inch art space. Several variations of a square were used to account for questions regarding edges and orientation; one with rounded corners, one turned as a diamond, and two with dimpled sides, straight and rounded, which increased perimeter edge. Both a normal pentagon and dimpled pentagon, or a star, were used. The final form was a 9-gon, the maximum single digit polygon and furthest from a circle. This set of ten polygons gives a cross section of basic form through which to test the effects of distortion, decreasing participant fatigue while still observing polygon typicality, or while still using forms considered basic polygons. Distortion was applied to these polygons in Adobe Illustrator, compressing or shrinking them horizontally, in the direction they would be lined up and also in the typical direction that graphics on FBSS labels are manipulated during the label application process.

All polygons were shown twice during the study, each being randomly matched with a high value of distortion, 6-10%, and a low value of distortion, 1-5%. The first polygon was selected using a number generator. Then the order was established by counting every third polygon until each was shown. Beginning with another shape and counting every third, a second round was established. Next, the percent distortion was assigned to the shapes by counting every third value, and making sure the second round of percent distortion was the difference between the first assignment and 10%. For example, if 8% were given to the circle, the next circle would be assigned 2% distortion. Finally, two control shapes with 0% distortion were randomly inserted into the sequence as placebo questions, one being a familiar and one an unfamiliar shape to preserve equality. These two placebos were used to see if there was any positional response bias for the participant pool. Thus, twenty-two (22) polygon questions embodied the survey, made using Microsoft PowerPoint. Blank white slides were used between each question to prevent learning and carry over effects, and an example with 20% distortion at the beginning demonstrated the experimental process. Among the polygons on each slide, the location of the distorted polygon could be in the left, middle, or right position. Each position was randomized so that, between slides, location never repeated more than twice consecutively and so that each of the three possibilities was equally probable. Each slide had three randomized letters in the top left, associated with the three polygons by same relative position. These were used to identify which polygons were unlike the other two. The stimulus changed but the task was consistent during the course of the survey. (Kanwisher et al., 1996; Orban, et al., 1996)

The same prompt was given before each participant took the survey, giving them the name, purpose, and instructions of the survey, and an opportunity was given for questions before proceeding. Every survey was administered on the same monitor in the same testing environment; lighting and distance were as consistent as possible. Follow up questions were asked regarding relevant demographics like gender, quality of eyesight, and experience with graphics. These factors could be considered in further analysis to see if there was any significance or correlation with detecting distortion.

***Results & Discussion***

Responses were collected and translated into correct versus incorrect identification of distorted polygon. Although there were 30 participants in the pilot study, which would invoke the Central Limit Theorem stating the data should be normally distributed, the correct/incorrect nature means the data is not normal but binomial. Data was analyzed using a Generalized Linear Model (GLM), which will be described in greater detail for the final study. Briefly, the GLM makes a linear function using the independent variables, distortion and relevant demographic data, to predict the dependent variable, the probability of detecting distortion. Given certain values for these independent variables, such as high distortion or high education, the probability of detection may increase above what is likely to occur by chance alone, and this function may be plotted to find the absolute threshold.

The accuracy of the linear function in this model is first analyzed in Table 1. The Rsquare value indicates the proportion of variation that can be attributed to the model rather than to random error. The GLM includes a Repeated Measures ANOVA (Table 2) that first indicates if any of the independent variables significantly influence the dependent variable. Because each degree of distortion was applied to two polygon types, the data for each polygon were combined so that distortion was consolidated by percentage. Again, ten polygons were used to decrease participant fatigue but differences between polygons were not an explicit concern of this study. The GLM compared the mean number of correctly identified distorted polygons across the ten degrees of distortion, 1-10%. The small value of Prob>F, 0.0001, indicates that the observed F-Ratio is unlikely, and is considered evidence that there is at least one significant effect in the model. Table 3 has a larger Prob>F, 0.761, confirming the model fits the data. Table 4 indicates that Percentage of Distortion is a significant factor influencing the response rate, but that Education and Age, while low Prob>F values, do not have significant effects but are worth noting. All other covariates were tested but had much higher Prob>F values and were not included here.

**Table 1: Pilot Summary of Fit**

| | |
|---|---|
| RSquare | 0.290211 |
| RSquare Adj | 0.273224 |
| Root Mean Square Error | 0.378796 |
| Mean of Response | 0.73 |
| Observations (or Sum Wgts) | 600 |

**Table 2: Pilot Analysis of Variance (ANOVA)**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 14 | 34.32032 | 2.45145 | 17.0849 |
| Error | 585 | 83.93968 | 0.14349 | **Prob > F** |
| C. Total | 599 | 118.26 | | <.0001* |

**Table 3: Pilot Lack Of Fit**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Lack Of Fit | 45 | 5.494167 | 0.122093 | 0.8405 |
| Pure Error | 540 | 78.445513 | 0.145269 | **Prob > F** |
| Total Error | 585 | 83.939679 | | 0.761 |
| | | | | **Max RSq** |

**Table 4: Pilot Effect Tests**

| Source | Nparm | DF | Sum of Squares | F Ratio | Prob > F |
|---|---|---|---|---|---|
| Distortion % | 9 | 9 | 33.36 | 25.8328 | <.0001* |
| Education | 3 | 3 | 0.737358 | 1.713 | 0.1632 |
| Age | 2 | 2 | 0.439881 | 1.5328 | 0.2168 |

*Distortion*

From the effects test, the significance was well under $P<0.05$, so it may be said with 95% confidence that correct response rate varied due to change in distortion, the independent variable. The null hypothesis is rejected, confirming the hypothesis; there is an absolute threshold at which a polygon is perceivably distorted and distinguishable from an original. While random polygon selection will result in 33% correct response rate, any percentage above may not necessarily be declared significant. Table 5 and Figure 9 illustrate the response rate according to percent distortion. Between 4 and 5% is where the correct responses begin to exceed 66%, halfway between the percentage found by guessing and always detecting distortion. Accounting for standard error and a level of certainty, the threshold under testing conditions is thus 5% distortion. It is important to note that the proportion of correct responses never quite reaches 100%. This confirms that the participant pool skews slightly right of normal distribution due to some individuals with low sensitivity or high indifference to the stimulus, distortion. Table 6 explicitly details which distortion percentages have means significantly different from each other. Average proportion of correct responses is included in both Tables 5 and 6.

**Table 5: Pilot Distortion LS-Means**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 1 | 0.24390491 | 0.06232883 | 0.25 |
| 2 | 0.42723825 | 0.06232883 | 0.43333 |
| 3 | 0.61057158 | 0.06232883 | 0.61667 |
| 4 | 0.61057158 | 0.06232883 | 0.61667 |
| 5 | 0.82723825 | 0.06232883 | 0.83333 |
| 6 | 0.74390491 | 0.06232883 | 0.75 |
| 7 | 0.91057158 | 0.06232883 | 0.91667 |
| 8 | 0.89390491 | 0.06232883 | 0.9 |
| 9 | 0.97723825 | 0.06232883 | 0.98333 |
| 10 | 0.99390491 | 0.06232883 | 1 |

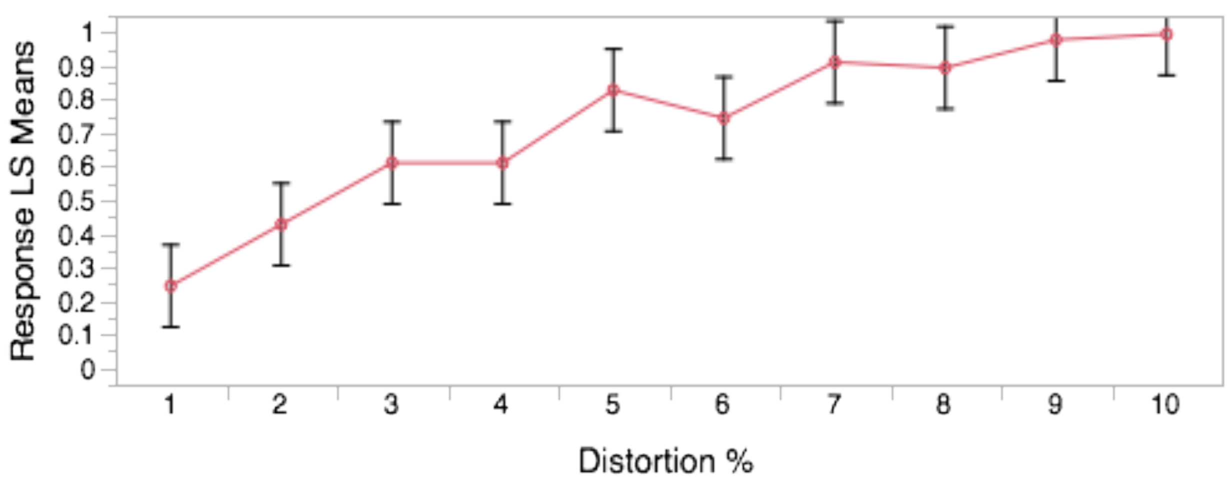


Figure 9: Pilot Distortion LS-Means Plot

**Table 6: Pilot Distortion LS-Means Differences Student's t-distribution**

α=0.050

t=1.96403

| Level | | | | | | | Least Sq Mean |
|---|---|---|---|---|---|---|---|
| 10 | A | | | | | | 0.99390491 |
| 9 | A | | | | | | 0.97723825 |
| 7 | A | B | | | | | 0.91057158 |
| 8 | A | B | | | | | 0.89390491 |
| 5 | | B | C | | | | 0.82723825 |
| 6 | | | C | D | | | 0.74390491 |
| 3 | | | | D | | | 0.61057158 |
| 4 | | | | D | | | 0.61057158 |
| 2 | | | | | E | | 0.42723825 |
| 1 | | | | | | F | 0.24390491 |

Levels not connected by same letter are significantly different.

*Education*

Education, with a low Prob>F value, does not have a significant effect but is worth noting and may be seen in Figure 10. Most participants had some college or a Bachelor’s Degree. With more testing, a slight trend could develop between highest education attained and mean correct response rate. Across all education levels, the proportion of correct responses was within range of threshold level. Standard error prevents a truly significant trend from being declared. Average proportion of correct responses is included with Table 7.

**Table 7: Pilot Education LS-Means**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 1 High School or Equivalent Degree | 0.62638889 | 0.09609178 | 0.65 |
| 2 Some College | 0.69561966 | 0.05109961 | 0.719231 |
| 3 Bachelor's Degree | 0.69722222 | 0.04752458 | 0.732143 |
| 4 Master's Degree | 0.87638889 | 0.06671389 | 0.825 |

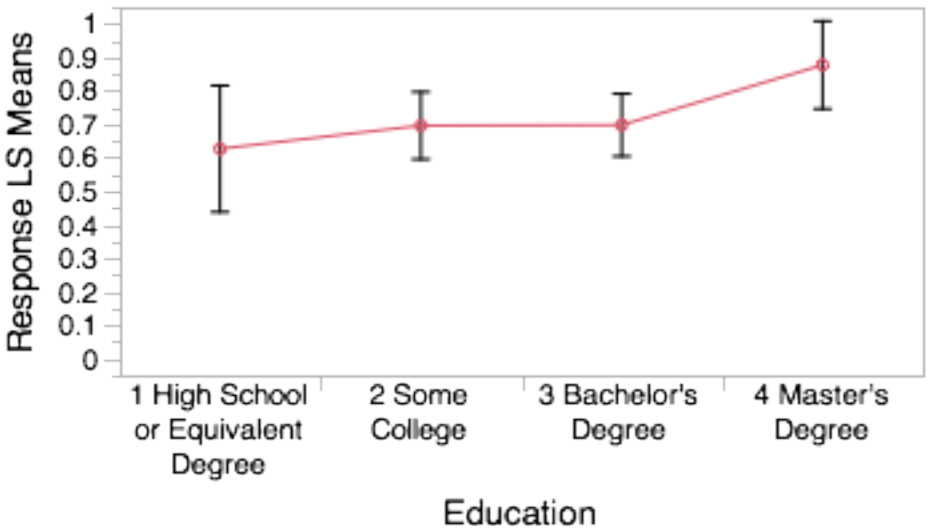


Figure 10: Pilot Education LS-Means Plot

*Age*

With a low Prob>F value, age also does not have a significant effect but is worth noting. In Figure 11, a slight trend appears as proportion of correct responses drops with increased age. Standard error again prevents declaration of a truly significant trend. Age typically impacts other factors such as education, eyesight, and experience with graphics, though none of these covariates were significant according to the GLM for this study. Average proportion of correct responses is included with Table 8.

**Table 8: Pilot Age LS-Means**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 18 to 24 | 0.74751603 | 0.03112314 | 0.724074 |
| 25 to 34 | 0.82668269 | 0.06967601 | 0.8 |
| 45 to 54 | 0.59751603 | 0.1083058 | 0.75 |

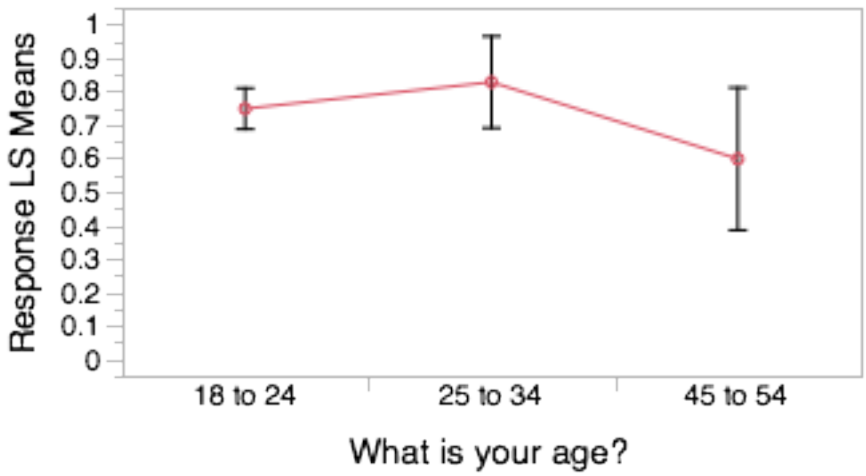

Figure 11: Pilot Age LS-Means Plot

*Placebo Response Location*

A chi-squared test was applied, comparing response frequency for the left, middle, and right locations, to determine whether people had response location bias. The two control questions with three identical polygons of 0% distortion were not counted as correct or incorrect, but position was noted for this purpose. With a P-value of 0.304, no significant response location bias was found, meaning participants were not inclined to answer in any certain way when unsure of the answer. Though not significant, there was a slight propensity for right of center, with fewer left and fewest middle polygons selected, as is seen in Table 9. If in future studies a positional bias was found, this could be compensated for by placing stimuli proportionally between locations or by formulating a different type of survey.

**Table 9: Pilot Chi-Square Means Table - Percent Spread**

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|
| Left | 208 | 31.5 | 31.5 | 31.5 |
| Middle | 186 | 28.2 | 28.2 | 59.7 |
| Right | 266 | 40.3 | 40.3 | 100 |
| Total | 660 | 100 | 100 | P-value = 0.304 |

***Conclusions***

Covariates included participant number, age, gender, ethnicity, education, eyesight, understanding of instructions, difficulty with survey, experience with graphics, and whether distortion would prevent a purchase decision. None of these yielded significance in effect. Data indicated only significance between degrees of distortion and response rate. Further testing with HVS capacity should examine if the threshold increases to higher distortion levels with more complex images, such as packaging labels. (Grill-Spector & Malach, 2004; Orban et al., 1996) Decreasing luminance contrast could also increase the difficulty of perceiving distortion and elevate threshold. (Pentland, 1988)

Limiting factors of this study include using only polygons, the standard black color, the brevity of the study, and the digital nature of the stimuli. The ability to show every polygon at every distortion level would be a longer though possibly more telling study. Also needed are greater participant numbers and wider demographics for more robust data. (Karunasekera & Kingsbury, 1995)

In spite of these limitations, it was found that in a context that maximized the probability and ease of perceiving distortion without encouraging false positives, the HVS has an absolute threshold of 5% at which a polygon is perceivably distorted and distinguishable from an original. There may also be factors that have not been considered affecting whether distortion is easily perceivable. Follow up studies should further investigate the nuances encompassed here but also determine if there is significance in effect between levels of familiarity, patterns, textures, typography, logos, faces, larger compositions, and photographs. (Isola et al., 2011; Orban et al., 1996) Other possible investigations include measuring the minimum amount of time required to identify distortion in an image (Thorpe, Fize, & Marlot, 1996), or using eye tracking technology to measure where and how people determine distortion.

# CHAPTER FOUR

# MATERIALS AND METHODS

The experiment "Estimating a 'consumer rejection threshold' for cork taint in white wine" by Prescott el al. (2005) was a prime foundation for the methodology of this final study. They combine a paired preference test with a method of constant stimuli threshold procedure. Their aim was to determine the consumer rejection threshold, seen in Figure 12, the point at which wine consumers would begin to reject a wine containing cork taint. Subjects were told that the study was a test of preference for different wines with no mention made of cork taint. (Prescott et al., 2005) However, in psychophysics situations where memory and sequence effects play a significant role, the 2-AFC requires greater sensitivity differences, resulting in slightly greater statistical power over the 3-AFC. The 2-AFC is generally used in such discrimination-based methods where the attribute change can be specified. (O'Mahony & Rousseau, 2002) For this final study, subjects were charged with a discrimination task for a specific quality, and were thus told to discern based on graphic distortion.

A pair of studies entitled "Product Design: Preference for Rounded versus Angular Design Elements" by Westerman et al. (2012) was also influential in the development of the final study. Through the first of these two computer-based studies, they "found a preference for rounded designs that extended to self-report purchase likelihood, with additive effects of contour and graphics shape that could not be accounted for by design typicality or perceived ease of use." Simultaneous manipulation of both contour shape and graphics shape leads to the emergent variable of "unity." Participants preferred designs with a high degree of unity, or congruence. The congruence associated with the combination of an angular contour and angular graphic will, to some degree, mitigate the generally negative effects of angularity on preferences.

Their second study, seen in Figure 13, was based on hedonic ratings of design qualities. It looked at the relationship between product (spring water vs. bleach), container shape (round vs. angular), and label shape (round vs. angular). They found that the interaction between bottle shape and graphics was not significant for any of the dependent measures, reinforcing that the effects of bottle shape and graphics shape are additive. Preference for design shapes was not product dependent. "When considering contour, people preferred designs that were more typical (the rounded designs) although their attention was captured by the atypical and less-preferred angular designs." (Westerman et al., 2012) While curved shapes were more typical for some products, preference for roundness did not vary. "Greater preference for rounded designs was apparent in participants' ratings of both aesthetics and purchase likelihood. Typicality is not a sufficient explanation of a relative preference for rounded designs. Perceived ease of use is also not a sufficient explanation of design preference." (Westerman et al., 2012) They also found independence of the effects of contour shape and graphics shape. If indeed there is independence, then relatively inexpensive promotional and seasonal changes can indeed be made to packaging graphics without needing to alter the shape of a product's container.

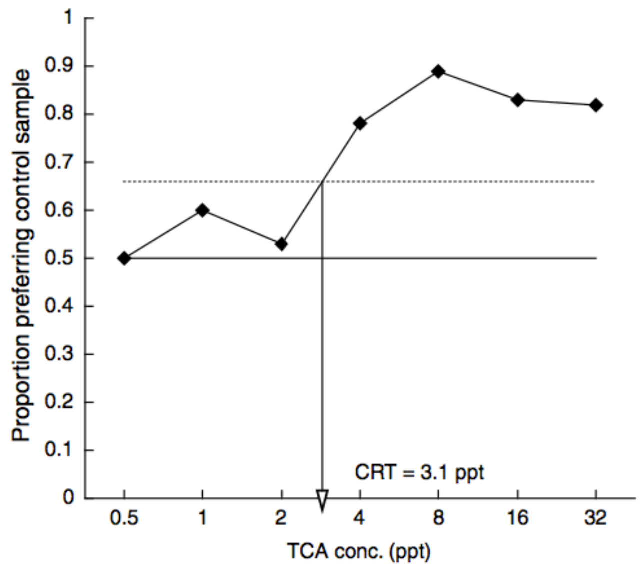


Figure 12. Proportion of subjects choosing wine without cork taint. Solid line represents no preference, while dotted line indicates 5% significance criterion threshold for N=58 subjects. (Prescott et al., 2005)

Figure 13. Experiment on the relationship between product, container shape, and label shape. (Westerman et al., 2012)

### *Objectives*

In the pilot study it was found that under optimal conditions subjects could perceive 5% or greater graphic distortion in basic polygons. This main study advanced the investigations of the pilot study to packaging labels, incorporating from the literature review information on FBSS, bottle technology and trends, the impact of product typicality, cultural understanding, and general label elements and context on the HVS. Because Bell et al. (2014) found that the HVS can interpret features together or independently, it was unclear what implications their findings would have for processing the more complex graphics of FBSS. This final study on graphic distortion in FBSS had treatments organized in a Randomized Complete Block Design (RCBD) and presented each of the stimulus levels along with a control for comparison using a 2-AFC Method. In this way, the present author hoped to answer the following research questions:

[RQ1]: Is there a measurable threshold of acceptable distortion?

[RQ2]: Does consumer preference for non-distorted graphics affect consumer acceptance of distortion?

[RQ3]: Does bottle shape affect consumer acceptance of distortion?

[RQ4]: Does familiarity with graphics affect consumer acceptance of distortion?

### *Experimental Design & Analysis*

The goal of this final study was to compare levels of graphic distortion on FBSS labels in order to find at which level consumers take notice. To compensate for the distortion that occurs while shrinking a label onto a bottle, software may be used to pre-distort graphics by varying amounts so a normal image appearance results after application. This process is described further in the “Stimuli” section. The first variable, distortion, had five levels: 0, 12.5, 25, 50, and 100 percent. Bottle shape was the second variable that was studied because of a bottle’s influence on the appearance of its label. Bottle shape had 3 levels: Concave, Convex, and Conical. The familiarity of a label was a third variable because of its potential impact on consumer attention to graphic distortion. Label familiarity had 2 levels: Brand Name and Generic. This experiment used a mixed 2 (labels) x 3 (bottle shapes) x 5 (distortion percentages) set up, effectively 30 treatments in a Randomized Complete Block Design (RCBD). “Label” was a between-subject variable, meaning each participant was tested with either the brand name or generic label, while bottle shape and distortion percentage were within-subject variables, meaning each participant was tested with all levels of bottle shape and all levels of distortion.

Often, more than one independent variable has an effect the dependent variable, and all factors should be in the model design and analysis. An extraneous independent factor is called a blocking factor, and while not the subject of the investigation its effect should be accounted for in the explanation of experimental results. The term “block” was first used by R.A. Fisher in 1926 while he was conducting agricultural field experiments. (Morris, 2011) Blocking divides experimental units into relatively homogeneous groups to make comparison between treatments more reasonable, decrease experimental error, and make results of inference more valid. In an RCBD, every block contains each treatment exactly once and treatments are assigned randomly within a block. Thus the blocks are randomized and complete. The RCBD is said to be an additive model, meaning there is no interaction effect between treatment and block.

Linear regression predicts the dependent response variable, Y, through a linear combination of independent observed variables, called predictors. This implies that a constant change in a predictor leads to a constant change in the response variable, which is appropriate when the response variable has a normal distribution. With RCBD, both blocks and treatments are the independent variables, and each data point of the dependent response variable is a function of the independent variables. The generalized linear model (GLM) is a flexible generalization of ordinary linear regression, allowing the response variable not to have a normal distribution and the variance of each measurement to be a function of its predicted value. The non-normal response variable is connected to the linear model through a link function that transforms the data. The GLM and SDT have been used in analysis of sensory discrimination tasks more recently than probabilistic Thurstonian models, mentioned previously.

The performance of a fixed number of trials with fixed probability of success on each trial is known as a Bernoulli trial. The Bernoulli distribution is the simplest discrete distribution, named after Swiss scientist Jacob Bernoulli. (McCulloch & Searle, 2001) When trying to predict the probability of success/failure or making a yes/no choice, a linear-response model is not appropriate. Data will not be normally distributed when probable outcomes are bounded as success, 1, and failure, 0. More simply, the probability itself is modeled rather than the frequency of observed outcomes. For this final study, each treatment had a Bernoulli distribution.

The statistical analysis here used a generalized linear mixed model (GLMM). Used widely in many disciplines, it is an extension of the GLM wherein the linear predictor contains random effects in addition to the usual fixed effects. By accommodating blocks in this linear mixed model, it is expected that any influence beyond fixed effects is random and independent from block to block. (Morris, 2011) These random effects are presumed to have a normal distribution. In the GLMM, fixed effects are used for modeling the mean of Y, the response variable, while random effects govern the variance-covariance structure of Y. A fixed effect is considered to be a constant, which may be estimated, while a random effect is an extraneous factor acknowledged to be affecting the data, which may be predicted. In this study, the fixed variables are Label, Bottle Shape, Distortion Percentage, and the random variable used for blocking is Person. Similar to the pilot study, stimulus levels were presented along with a control for comparison using the 2-AFC Method. Presentation was changed from 3-AFC Method to more closely align with a choice of preference and a Bernoulli trial.

In all experiments, the basic GLM equation is:

$$Y_{ij} = \mu + r_i + \beta_j + \varepsilon_{ij}$$

| | |
|---|---|
| $Y_{ij}$ | response to treatment i in block j |
| $\mu$ | overall average response to all treatments |
| $r_i$ | effect of treatment i |
| $\beta_j$ | effect of block j |
| $\varepsilon_{ij}$ | random error associated with treatment i in block j |

For this final experiment, the specific GLMM equation is:

$$Y = \mu + L + S + D + (S*D) + (L*S) + (L*D) + (L*S*D) + P(L) + \varepsilon$$

| | |
|---|---|
| $Y$ | response to a given treatment in a specific block |
| $\mu$ | overall average response to all treatments |
| $L$ | fixed effect of Label (brand name vs generic) |
| $S$ | fixed effect of Bottle Shape (concave, convex, conical) |
| $D$ | fixed effect of Distortion % (0, 12.5, 25, 50, 100) |
| $(S*D)$ | interaction effect between Bottle Shape and Distortion % |
| $(L*S)$ | interaction effect between Label and Bottle Shape |
| $(L*D)$ | interaction effect between Label and Distortion % |
| $(L*S*D)$ | interaction effect between Label, Bottle Shape, Distortion % |
| $P(L)$ | random effect of blocking by Person (within each label study) |
| $\varepsilon$ | random error associated with a given treatment in a specific block |

***Hypothesis***

[RQ1]: Is there a measurable threshold of acceptable distortion?

Any percentage of distortion that produces a proportion of correct responses equal to or greater than 75% is considered the threshold according to classical psychophysical threshold testing measures. (Gescheider, 1997; ASTM E1808)

$H_0$: $\mu_d$ <0.75, where d is some amount of distortion

$H_a$: $\mu_d$ ≥0.75, where d is some amount of distortion

Using the GLMM method, compare fixed effects of distortion percentage on proportion of correct responses.

$H_0$: $\mu_0=\mu_{12.5}=\mu_{25}=\mu_{50}=\mu_{100}$

$H_a$: $\mu_0\neq\mu_{12.5}\neq\mu_{25}\neq\mu_{50}\neq\mu_{100}$

[RQ2]: Does consumer preference for non-distorted graphics affect consumer acceptance of distortion?

Compare proportion of correct responses of those who self-reported that distortion would or would not affect their decision to purchase the product.

$H_0$: $\mu_{Yes}=\mu_{No}$

$H_a$: $\mu_{Yes}\neq\mu_{No}$

[RQ3]: Does bottle shape affect consumer acceptance of distortion?

Compare fixed effects of bottle shape on proportion of correct responses.

$H_0$: $\mu_{Convex}=\mu_{Concave}=\mu_{Conical}$

$H_a$: $\mu_{Convex}\neq\mu_{Concave}\neq\mu_{Conical}$

[RQ4]: Does familiarity with graphics affect consumer acceptance of distortion?

Compare fixed effects of label on proportion of correct responses.

$H_0$: $\mu_{BrandName}=\mu_{Generic}$

$H_a$: $\mu_{BrandName}\neq\mu_{Generic}$

***Stimuli***

In order to generate the digital stimuli for graphic distortion, 3D bottles were first modeled in Solidworks. Standard top and bottom pieces were made and a height established. Then, three forms for the middle of the bottle were modeled, based on common bottle types observed in retail environments, including a concave, convex and conical shape. These middle sections varied from the standard top and bottom by circumference, because distortion occurs where a FBSS shrinks around a smaller circumference on a bottle. To standardize this difference for each bottle shape, the ratio in circumferences roughly followed Fechner's golden ratio (1:1.7, approximates 1:1.618). For the concave bottle, the circumference of the top and bottom was 1.7 that of the middle. For the convex bottle, the middle was 1.7 that of the top and bottom. For the conical bottle, the top circumference was reduced so that the taper followed the height of the bottle and so the bottom was 1.7 that of the neck's circumference. These three bottle shapes were saved as basic .DAE (Solidworks) files, seen below in Figure 14.

Figure 14. Bottles created in Solidworks.

The Solidworks bottle files were then opened in Esko's Shrink Sleeve Toolkit, which simulates FBSS application and resulting geometry. For this process, the maximum accuracy setting was selected, as opposed to speed, and the standard settings for material properties (10% vertical, 65% horizontal shrinkage) and coefficient of friction (0.35) were used. All sleeve were shrunk for a duration of 6 seconds of simulation time, long enough for all transformation to occur. The 3D bottle with simulated sleeve geometry applied was then saved as a .ZAE (Collada) file. To simulate the graphic distortion of labels, the Collada files were brought into Adobe Illustrator. Here, a template was opened showing the layflat artspace of the FBSS prior to application. Two separate labels were made, a custom-made generic yogurt drink label, Figure 15, and a brand name coffee product label, Figure 16. This was to simulate label familiarity and test whether sensitivity to distortion increases with familiarity, as might be expected with leading brand name labels. Product typicality was maintained since both products, yogurt and coffee, were among those observed utilizing the expanding number of contoured bottle shapes and FBSS. Graphics created beforehand were dropped into the layflat artspace and simulated onto the bottle in a separate window via the previously generated sleeve geometry. Unique .AI (Illustrator) files were saved for each bottle type, containing the bottle, shrunken FBSS geometry, and graphics.

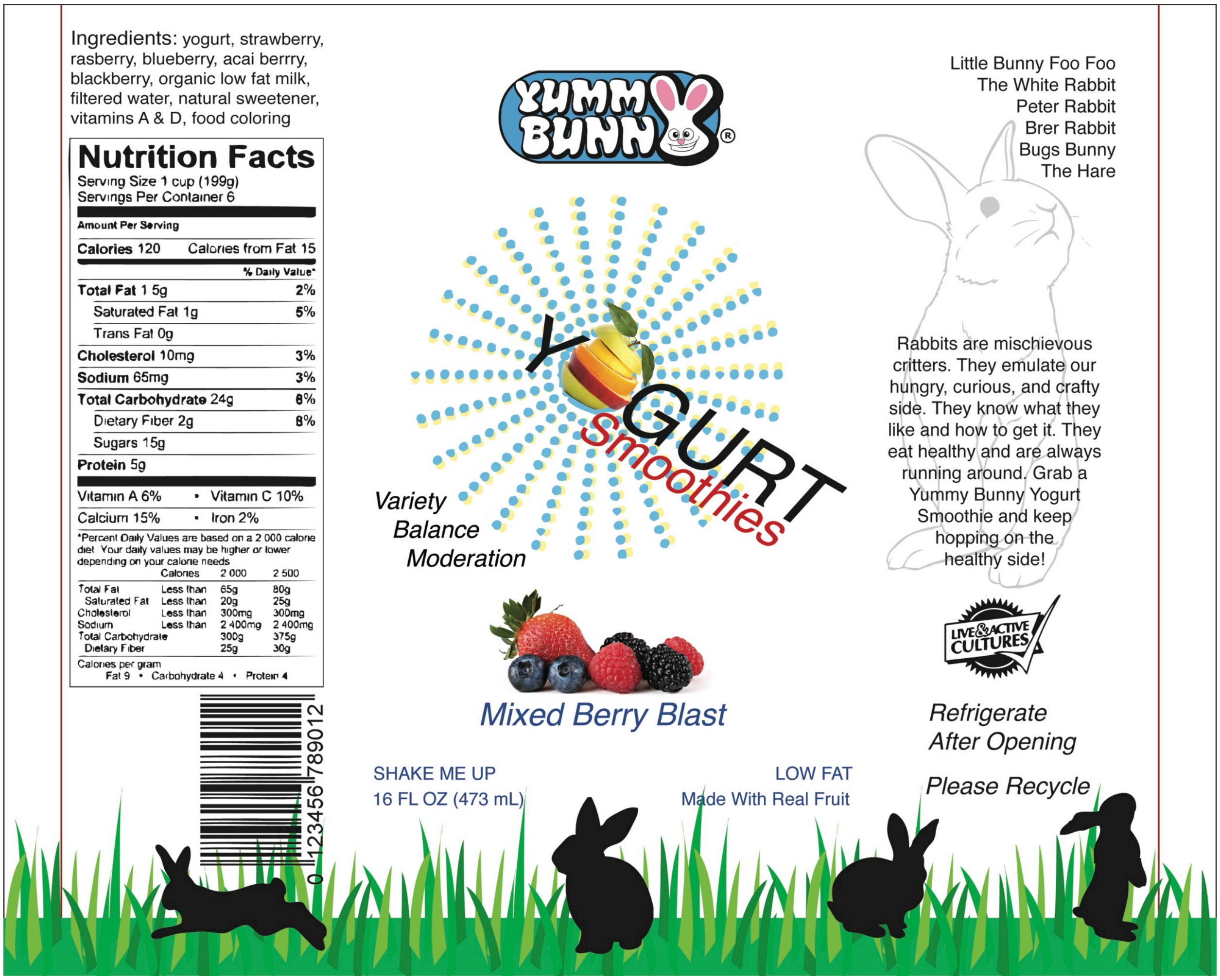


Figure 15. Generic label, Yummy Bunny Yogurt.

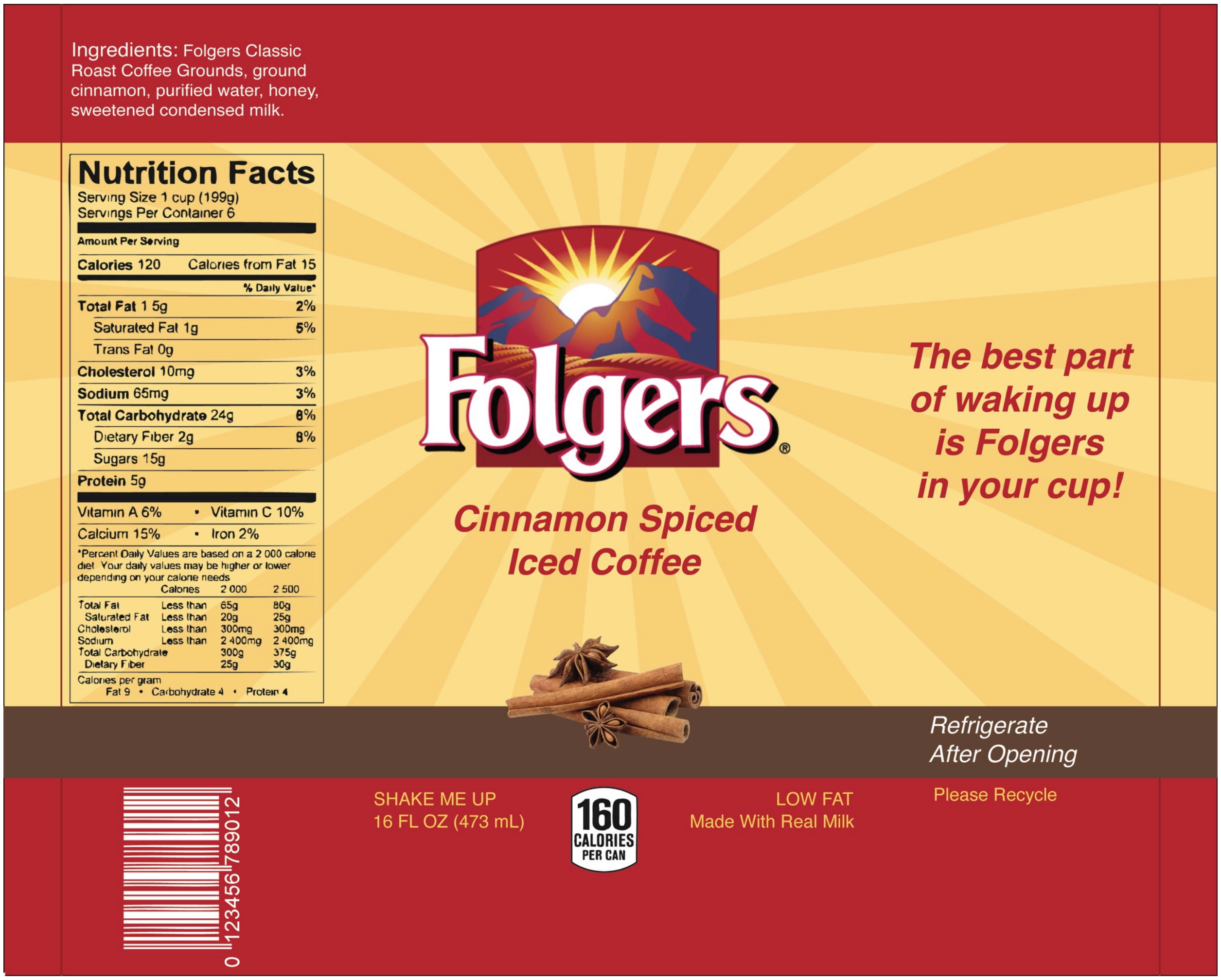


Figure 16. Brand name label, Folgers Coffee.

Distortion was varied during the next step using Esko's Shrink Sleeve Toolkit. A pre-distort feature allows graphics to be manipulated in preparation for application to the bottle. This feature has two options; style 1 allows bottle shape to determine distortion, and style 2 simulates a flat projection from a single perspective. Style 1 was utilized for this study since bottle shape certainly impacts graphics, and because style 2 only works from a single perspective, which is not maintained in an actual shopping context where tangible packaged goods may be held and moved. All graphics on both labels were grouped into four regions, each roughly a quadrant of label height, to locally address distortion since circumference varies along bottle height. Graphics were pre-distorted in place on the label using a sliding tool to set to a percentage. Since Fechner's Law states the relationship between sensory experience and stimulus intensity is nonlinear, each increment was double the previous one so the progression of graphic distortion went 0, 12.5, 25, 50, and 100 percent. (Sage & Goldstein, 2010) Another way of saying this is that using the slider tool, pre-distortion decreased from 100, 87.5, 75, 50, and 0 percent.

Within Adobe Illustrator, a visualizer window allows one to see how the pre-distorted label will appear on the bottle after the simulated shrink process. In this window, options were adjusted such as the label material (glossy opaque white PE or OPP film), the scene/environment (Supermarket Food), and the preset view, which was established so that the bottle would appear directly ahead, as it would on a shelf at eye level for a shopper of average height. PNG images of the bottle with label applied were then produced for each increment of distortion. These original PNGs were 21.3 inches wide by 12.3 inches tall with 100 dots/inch in resolution. Art optimized anti-aliasing was also applied so that any pixelated graphics were smoothed out through blending pixel colors.

Using Illustrator again, these images were compiled into pairs, a bottle with 0% distortion (control) and the same bottle shape with an increment of distortion (stimuli). Each pair was a set distance apart, and saved together using the artboard space as a new comparison image. The position of the control bottle was randomly selected each time.

Each comparison was to be a slide in the experimental survey. These comparison images were uploaded into SurveyMonkey online survey software. The sequence of comparison images was also set to randomize for every participant. Two surveys were made, one for the generic label and one for the brand name label, each comparing all levels of distortion to a control for each of the three bottle shapes. All bottles with brand name coffee labels and with generic yogurt labels may be seen in Figure 17 and Figure 18, respectively.

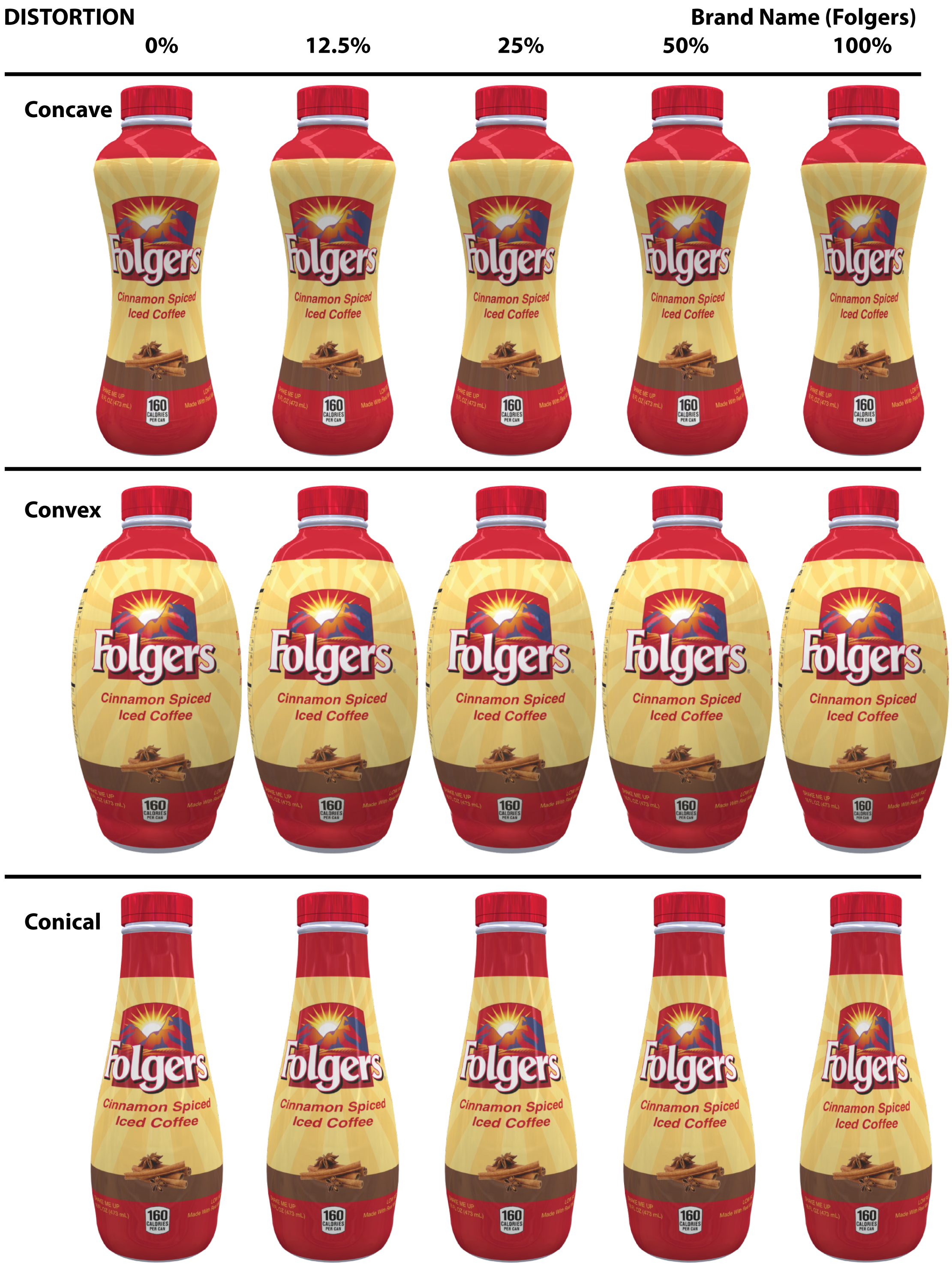


Figure 17. Comparing distortion percentages with brand name label on bottles.

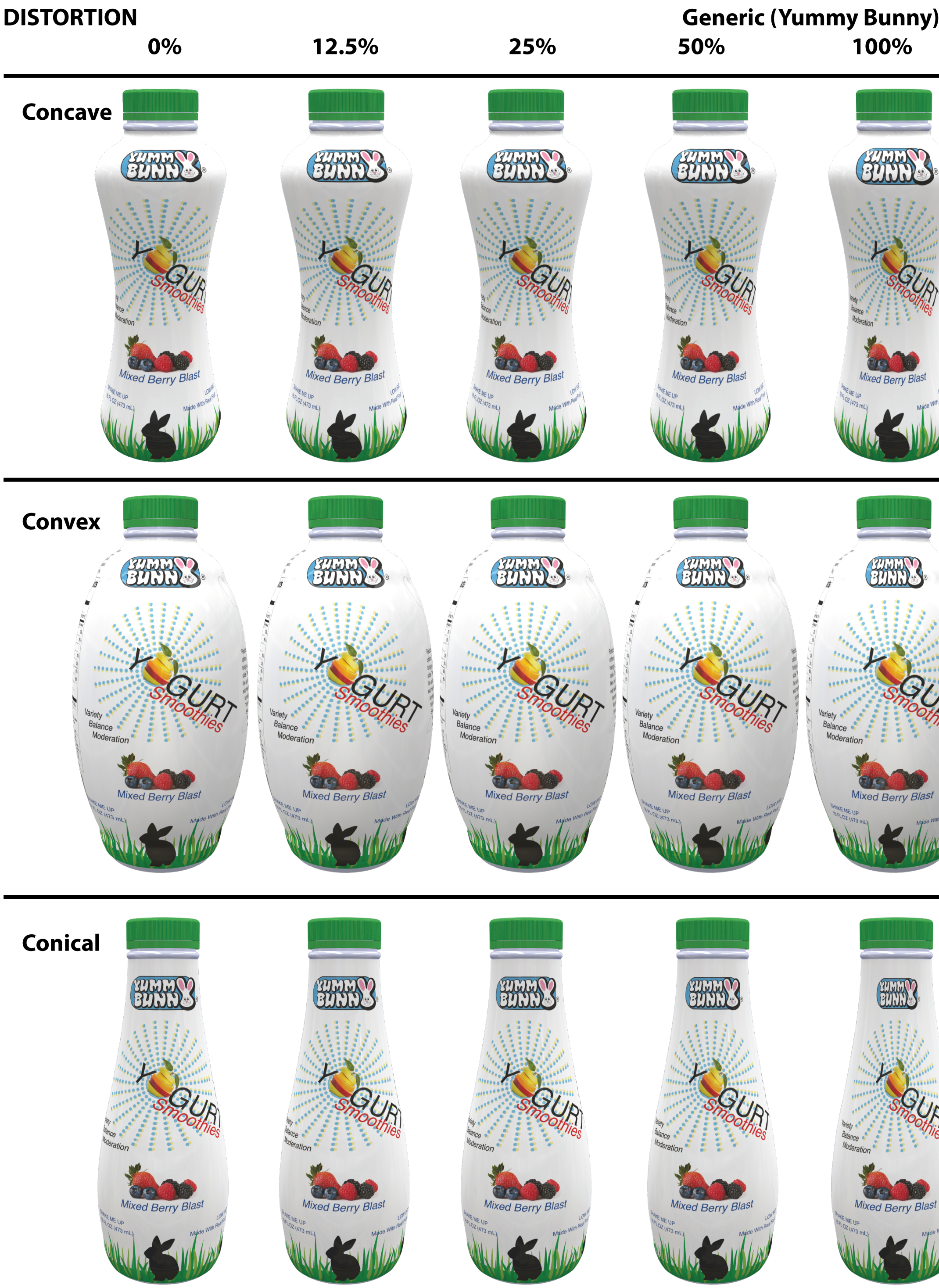


Figure 18. Comparing distortion percentages with generic label on bottles.

***Experimental Procedure***

These two surveys were then published online using Amazon's Mechanical Turk (MTurk) software. More about MTurk and its soundness for academic research is available in Appendix A. Because the surveys were online, some testing conditions could not be strictly enforced. More about the comparability of screen size is available in Appendix B. It is assumed that the surveys were taken indoors where most computers are housed and used, limiting testing conditions to some extent. Information about the nature of the experiment was provided prior to a participant beginning voluntarily. Instructions for the survey were also provided as a participant progressed along. The stimuli presentations were followed by demographic questions and the experiment concluded with a completion code, taken from SurveyMonkey back to the prompt in MTurk for verification and compensation. Each participant received $0.25 for completing the survey. The maximum time allotted for taking the survey was one hour. Data was collected over a period of five days in January 2015.

# CHAPTER FIVE

# RESULTS AND DISCUSSION

***General Demographics***

In all, there were 200 participants, 100 per survey. 70 were 20-30 years old, 69 were 30-40 years old, 32 were 40-50 years old, 23 were 50-60 years old, and 6 were over 60 years old. None were under 20 years old. Of this group of 200, 94 were male and 106 were female. 153 were Caucasian, 11 were African-American, 10 were Hispanic, 22 were Asian / Pacific Islander, and 1 was American Indian. 6 participants had completed Some High School, 25 High School or Equivalent Degree, 57 Some College, 20 Associate Degree, 74 Bachelor's Degree, 16 Master's Degree, and 2 had completed a Doctorate Degree. 178 participants had 20/20 vision or vision corrected to 20/20, 14 did not, and 8 did not know. On a scale of 1-5 for understanding the task of the study, 1 being not very well and 5 being very well, the mean rating of understanding was 4.3. On a scale of 1-5 for difficulty finding distortion during the study, 1 being not very difficult and 5 being very difficult, the mean rating of difficulty was 3.8. Considering who had education or experience working with graphic design, illustration, etc., 34 did while 166 did not. With respect to education or experience working with shrink-wrap film or labels, 11 did and 189 did not. Considering all bottles in the survey and if any of the differences in graphics would change their decision to purchase the product, 25 said "yes" and 175 said "no."

### *Outliers Excluded*

Data did include some outliers, which were excluded from analysis based on criteria of time, response rate, and self-reported understanding. There were 17 unique participants excluded from the final analysis of 183 participants.

The first qualification was any participant that took less than 2 minutes to complete all of the 15 experimental questions and 10 required demographic questions. This is roughly the minimum time required to complete all questions after quickly reading instructions. 3 participants were excluded on this basis.

Another qualification was response rate, or total correct responses. The probability of a participant getting correct 2 or fewer of the 12 experimental questions containing stimuli is very low, even if participants were guessing (With 12 trials, binomial probability $P(x=2)$: 0.01611328, $P(x=1)$: 0.00292969, $P(x=0)$: 0.00024414, and $P(x=0,1, \text{ or } 2)$: 0.01928711 cumulative). It is more likely that they mixed up instructions of the study. 7 participants were excluded on this basis.

No participants were excluded on the basis of getting more correct as this would be attributable to the stimuli. With 11 as the most, no participant correctly responded on all 12 experimental questions.

The final qualification was self-reported understanding of the task, with a scale from “1 - not very well” to “5 - very well.” Those who reported understanding at 1 or 2 were excluded because they may have performed the study with the wrong intent, such as selecting labels that were not distorted rather than those that were distorted. 7 participants were excluded on this basis.

***Observations from Demographics***

JMP is a business unit of SAS that produces interactive software for desktop statistical discovery. Pronounced "jump," its name suggests a leap in interactivity, a move in a new direction. JMP was used for most of the statistical analysis on demographic information and graphing.

**Table 10: Effect Tests**

| Source | Nparm | DF | Sum of Squares | F Ratio | Prob > F |
|---|---|---|---|---|---|
| Gender | 1 | 1 | 1.8340546 | 7.4253 | 0.0065 |
| Age | 4 | 4 | 1.7614198 | 1.7828 | 0.1295 |
| Ethnicity | 5 | 5 | 1.6033249 | 1.2982 | 0.2618 |
| Education | 6 | 6 | 2.4093241 | 1.6257 | 0.1360 |
| Decision to Purchase? | 1 | 1 | 0.1129550 | 0.4573 | 0.4990 |
| Graphic Experience | 1 | 1 | 0.0441771 | 0.1789 | 0.6724 |
| Shrink Wrap Film Experience | 1 | 1 | 0.5626646 | 2.2780 | 0.1314 |
| 20/20 Vision | 2 | 2 | 0.6440728 | 1.3038 | 0.2717 |
| Time | 1 | 1 | 0.0197760 | 0.0801 | 0.7772 |
| Difficulty | 1 | 1 | 0.1223067 | 0.4952 | 0.4817 |
| Understanding | 1 | 1 | 0.0112615 | 0.0456 | 0.8309 |

Table 10 shows that the only demographic variable that had a significant fixed effect was Gender. In Table 11, the average man had a correct response rate 0.074 higher, or 7.4% better than the average woman.

**Table 11: Gender Response Rate LS_Means**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| Female | 0.53758206 | 0.05540012 | 0.486979 |
| Male | 0.60314370 | 0.05583903 | 0.561303 |

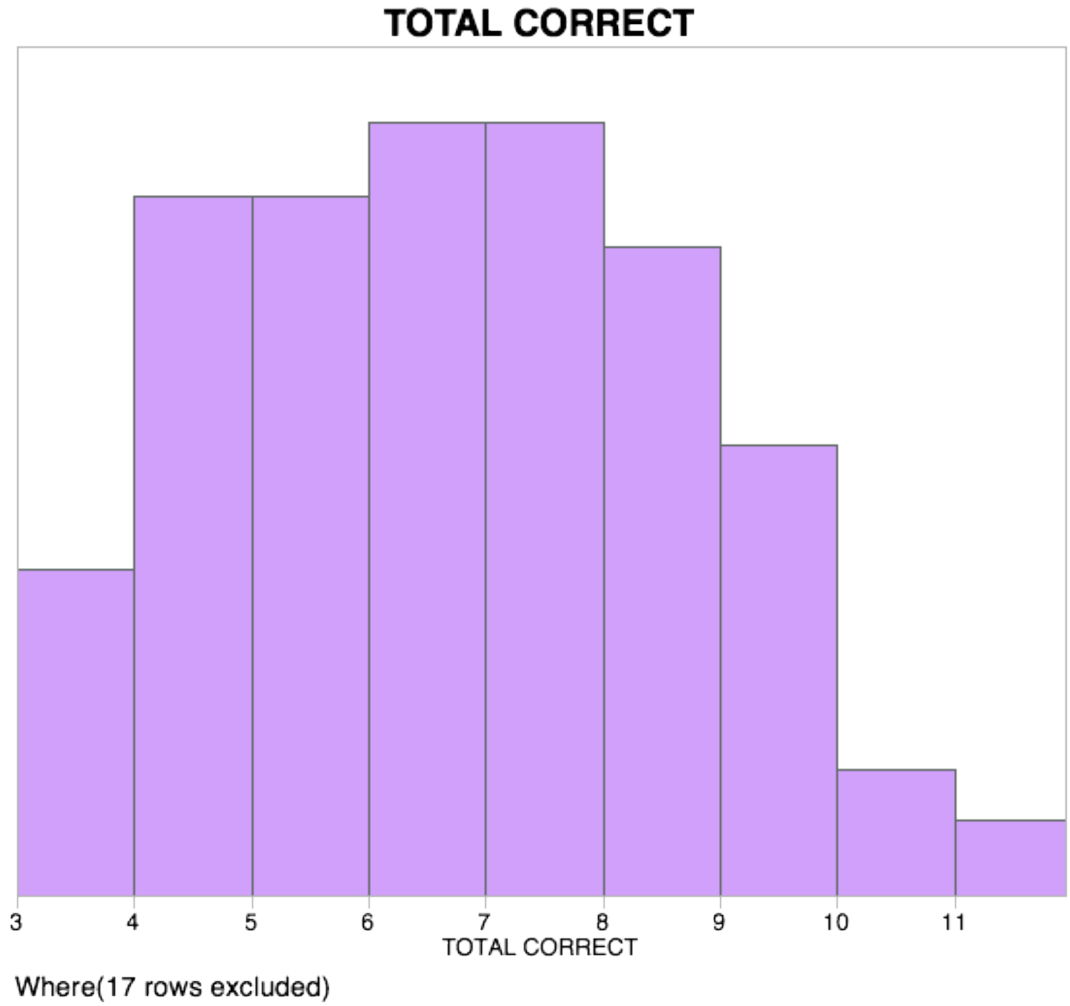


Figure 19. Number of Participants Correctly Identifying Distorted Sample.

Seen in Figure 19, the participant scores for correctly identifying distorted labels are roughly normally distributed, with a mean just above 6 labels. Skewed right just slightly, this indicates that very few participants could detect nearly all percentages of distortion, and that the lowest levels of distortion were rarely identified.

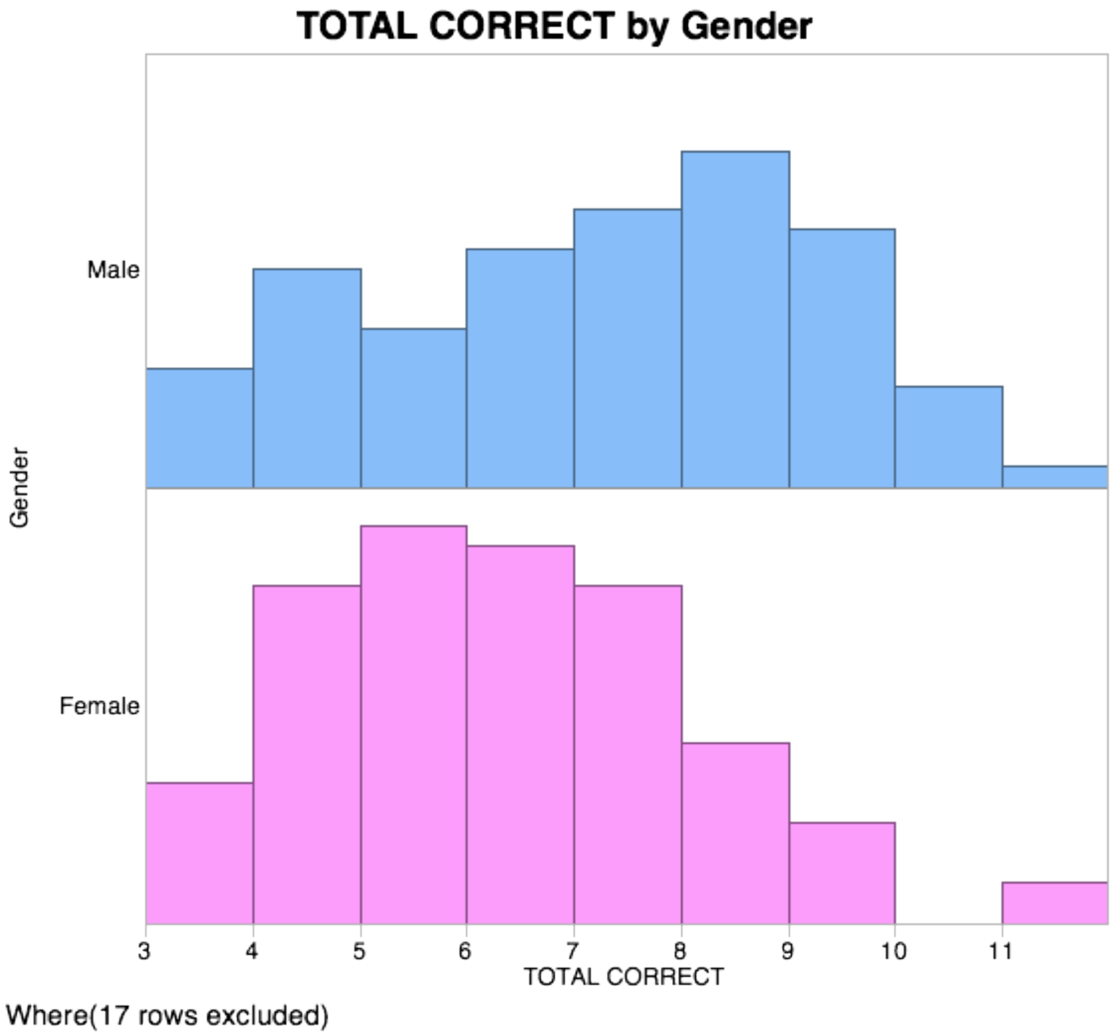


Figure 20. Number of Participants Correctly Identifying Distorted Sample by Gender.

**Table 12: Gender Total Correct LS-Means**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| Female | 6.4509847 | 0.72426263 | 5.84375 |
| Male | 7.2377244 | 0.73000062 | 6.73563 |

The distribution of the overall histogram is revealed to be bimodal when sorted by gender in Figure 20. In Table 12, it is seen that men scored higher with a mean of nearly 7, and skewed left indicating not all men had an advantage. Women had a mean of nearly 6, and skewed right indicating some women performed much better.

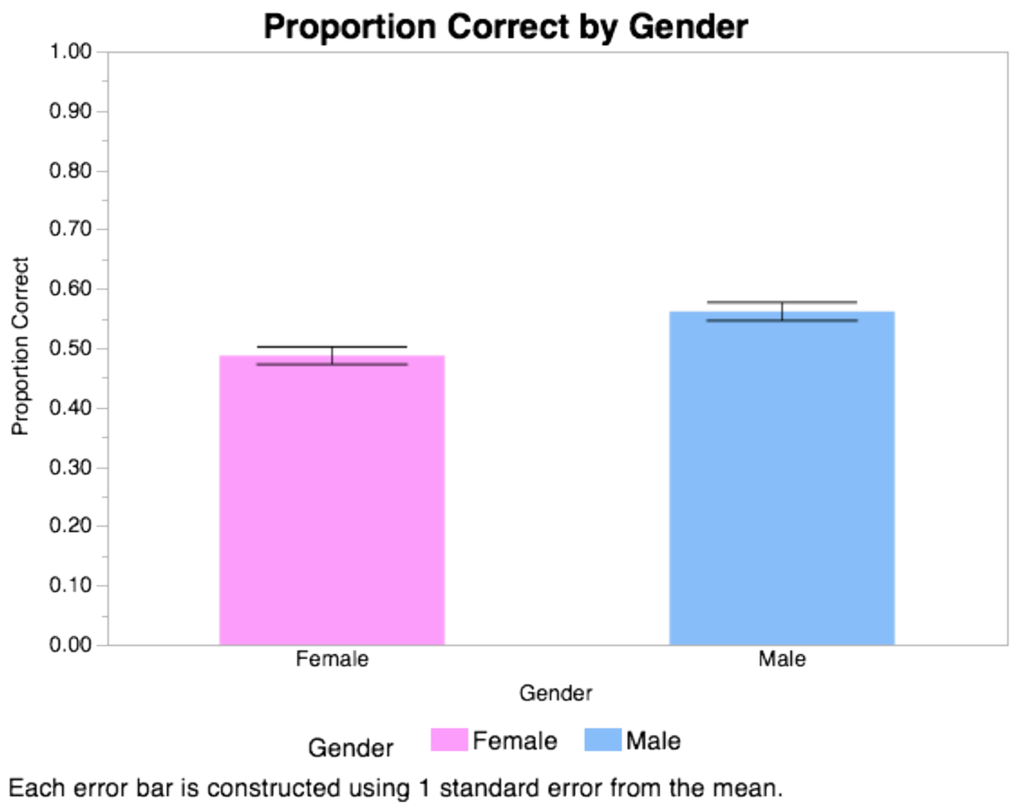


Figure 21. Bar Graph of Proportion Correct by Gender

Men identified distortion significantly better, as seen in Figure 21. This is possibly because they tend to be more visually oriented, more image-driven in their preferences for specific designs, and more concerned with overall product structure, whereas women pay more attention to details, textures and natural-looking forms. Men view design simplicity in image-related terms, while women prefer it in practical terms. (Xue & Yen, 2007) On average, women did no better than guessing with 50% correct, while men had close to 60% correct. Rather than being more perceptive, it may be that men were more competitive and had more personal investment in getting as many correct as possible, while women were simply finishing the survey for compensation.

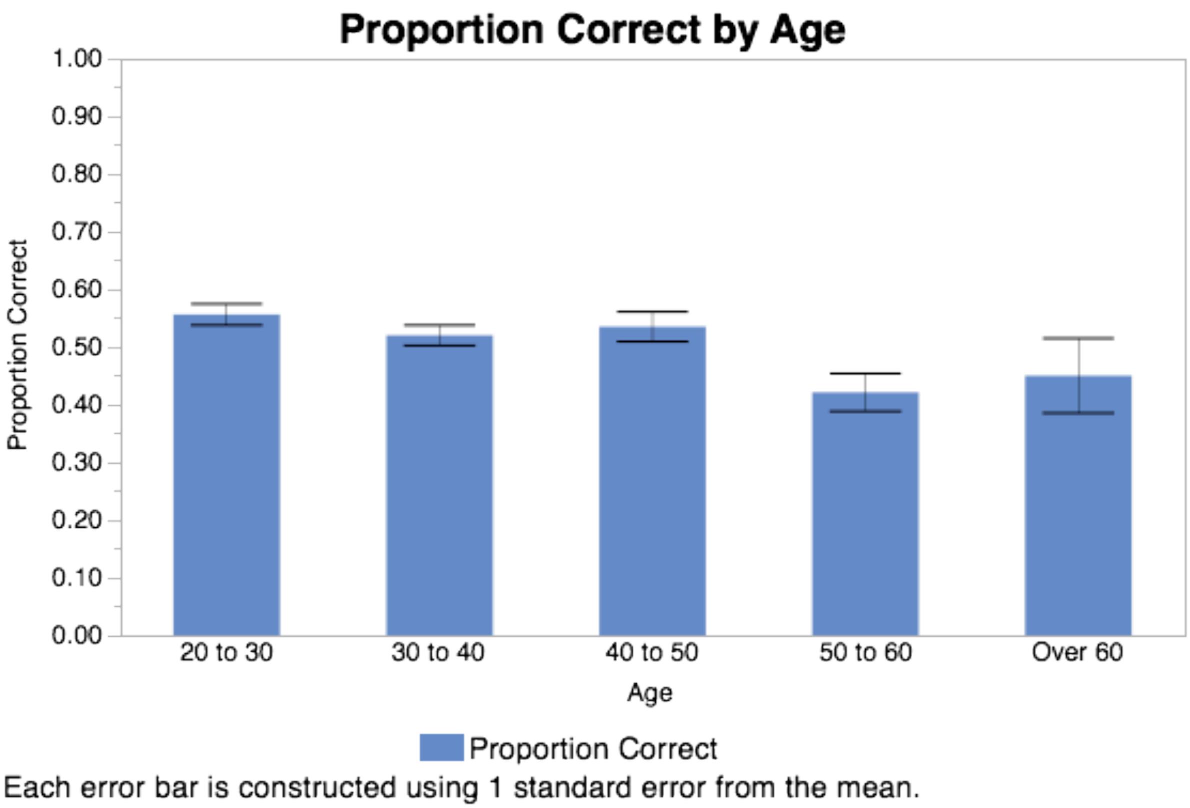


Figure 22. Bar Graph of Proportion Correct by Age.

**Table 13: Age LS-Means**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 20 to 30 | 0.62125075 | 0.05215583 | 0.556011 |
| 30 to 40 | 0.59339292 | 0.05550690 | 0.519900 |
| 40 to 50 | 0.61276616 | 0.05884464 | 0.534946 |
| 50 to 60 | 0.52962048 | 0.06272961 | 0.421053 |
| Over 60 | 0.49478410 | 0.08735626 | 0.450000 |

Older ages appear to score lower in Figure 22, but age was not significant. Mean response rate by age is detailed in Table 13. Age may be considered in future studies because of its impact on other variables such as technological adeptness and vision.

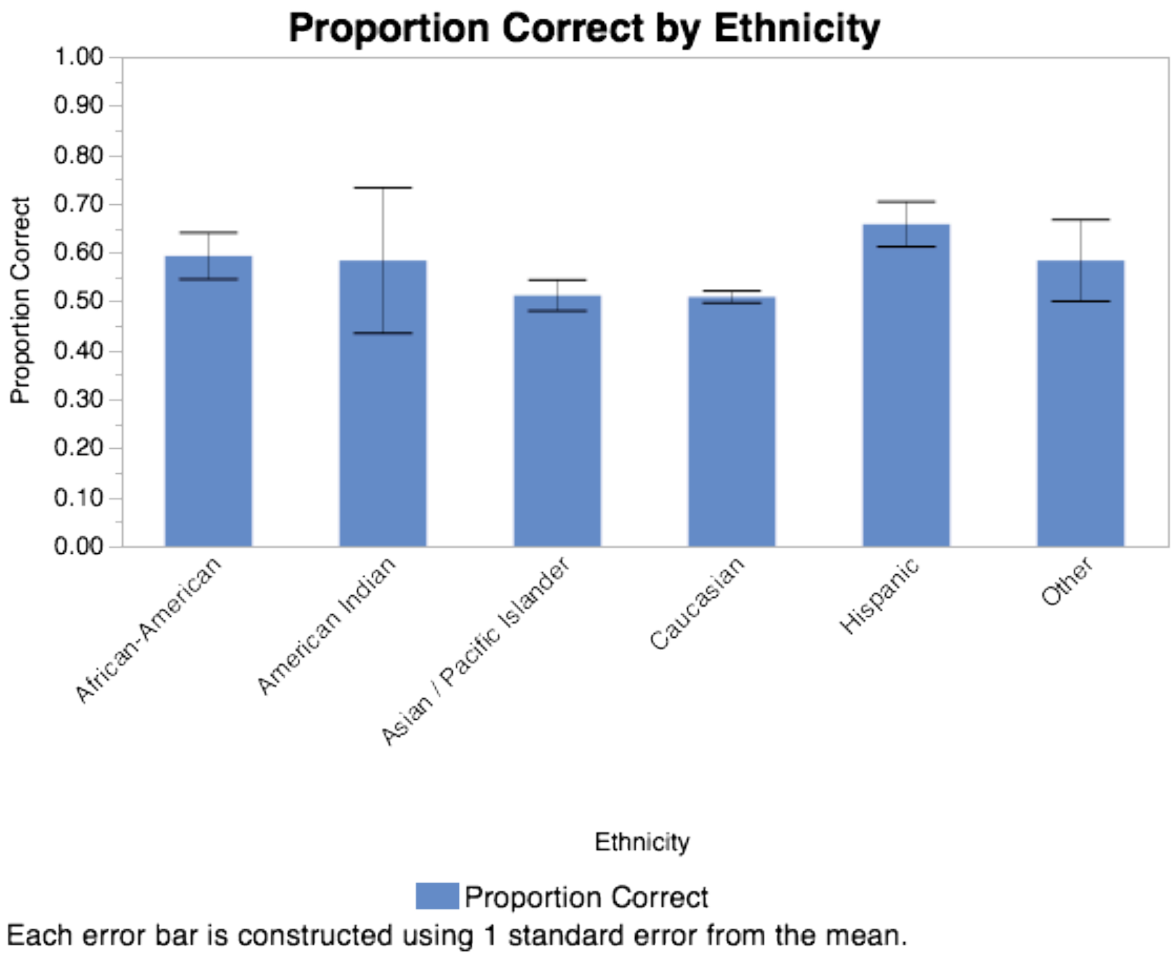


Figure 23. Bar Graph of Proportion Correct by Ethnicity.

No significance was found for ethnicity. See Figure 23.

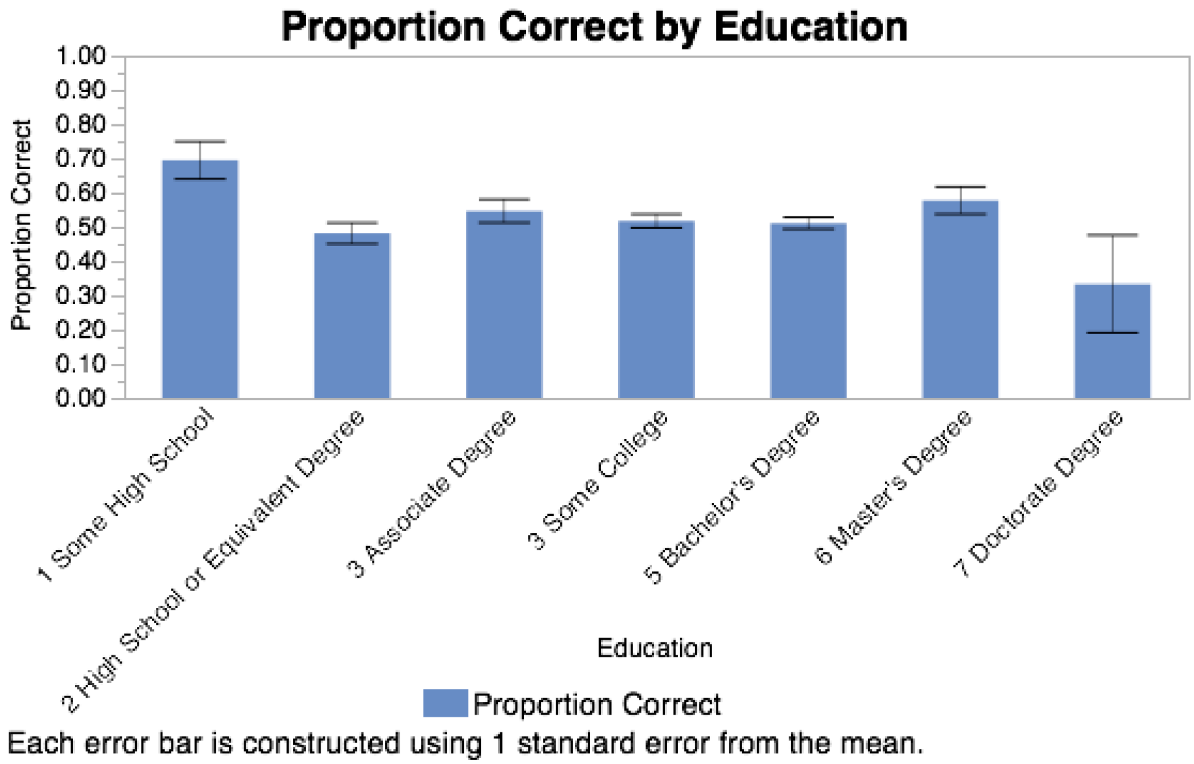


Figure 24. Bar Graph of Proportion Correct by Education.

**Table 14: Education LS-Means**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 1 Some High School | 0.68931895 | 0.08664465 | 0.694444 |
| 2 High School or Equivalent Degree | 0.51798019 | 0.05958054 | 0.481061 |
| 3 Associate Degree | 0.61818034 | 0.06217470 | 0.546296 |
| 3 Some College | 0.56071871 | 0.05174233 | 0.517296 |
| 5 Bachelor's Degree | 0.56611781 | 0.05243647 | 0.510714 |
| 6 Master's Degree | 0.62453577 | 0.06007362 | 0.576923 |
| 7 Doctorate Degree | 0.41568840 | 0.15495072 | 0.333333 |

Education did not have a significant effect, though in Figure 24 participants with only some high school education appeared to do slightly better and those with a doctorate degree slightly worse. This may be due to other factors, such as age and eyesight. Younger participants may also be more technological adept or have free time to participate in online studies, as very few participants with doctorate degrees participated. Participation in MTurk surveys may also be a primary source of income for those in high school, likely not the case for those with further education. Mean response rate by education is detailed in Table 14.

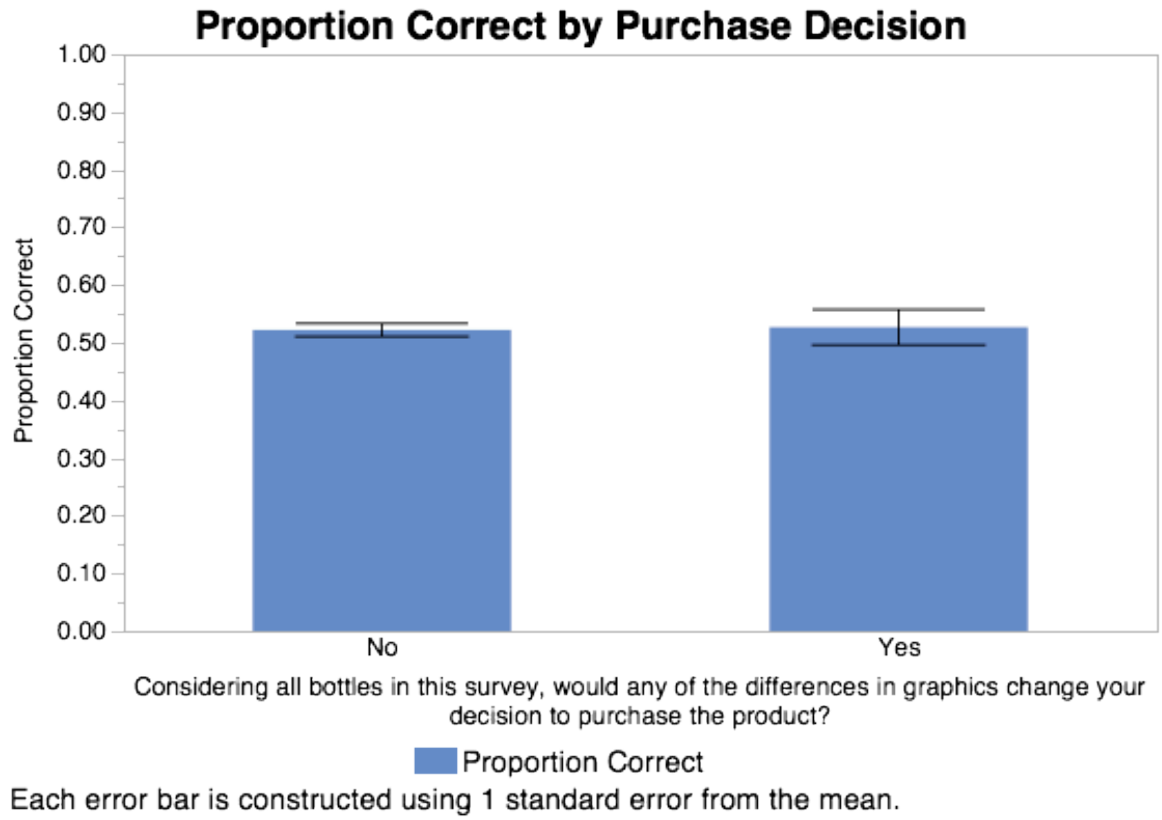


Figure 25. Bar Graph of Proportion Correct by whether distortion would impact a decision to purchase the product (self-reported).

Even when participants self-reported that distortion would impact their purchase decisions, their sensitivity to distortion remained the same, seen in Figure 25. No significance was found for having a preference. If mindful consumers cannot detect distortion with any greater ease, their preferences have no impact. Of those 183 participants included in analysis, 161 reported “no” and 22 reported “yes.”

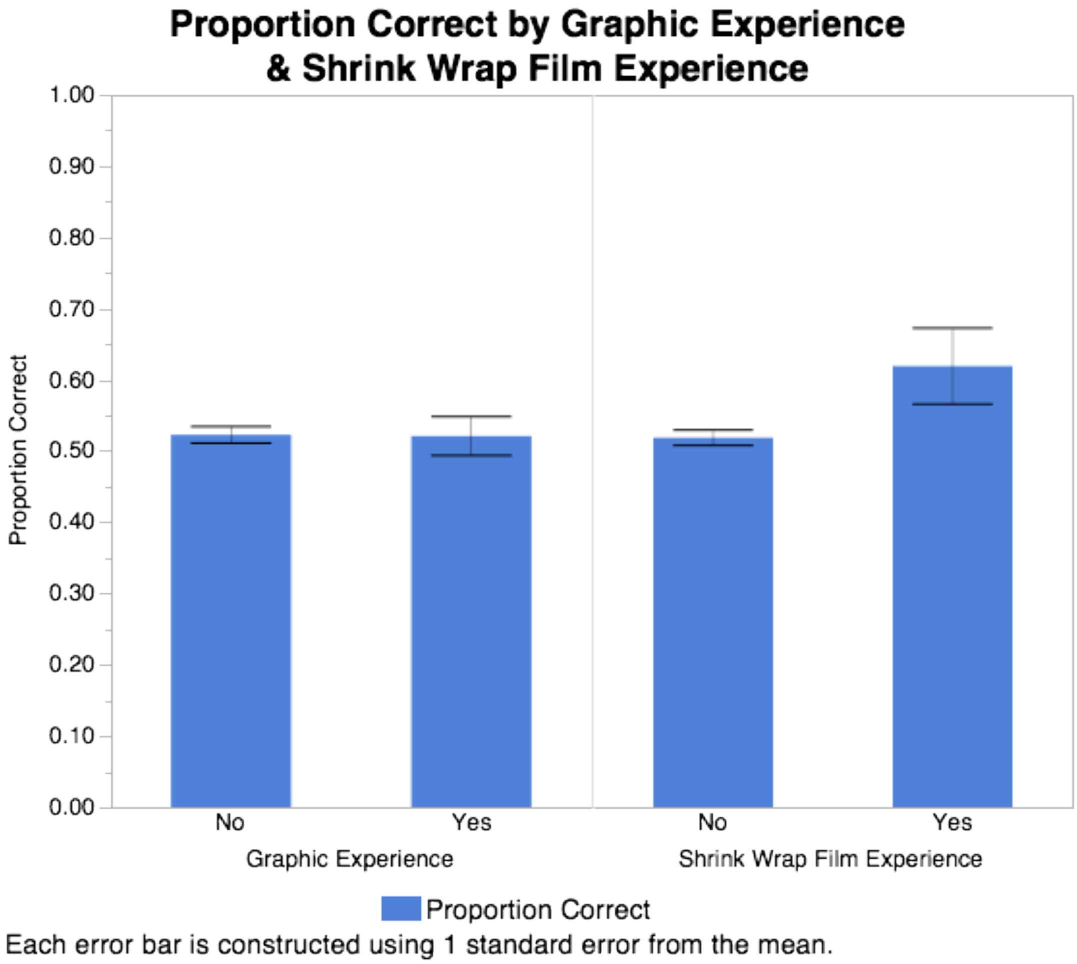


Figure 26. Bar Graph of Proportion Correct by Graphic Experience & Shrink Wrap Film Experience.

**Table 15: Shrink Wrap Film Experience LS-Means**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| No | 0.52614989 | 0.04770147 | 0.518466 |
| Yes | 0.61457587 | 0.07306530 | 0.619048 |

In Figure 26, participants with graphic experience did no better than those without, and those with shrink-wrap film experience did no better than those without. Mean response rate by shrink-wrap film experience is detailed in Table 15.

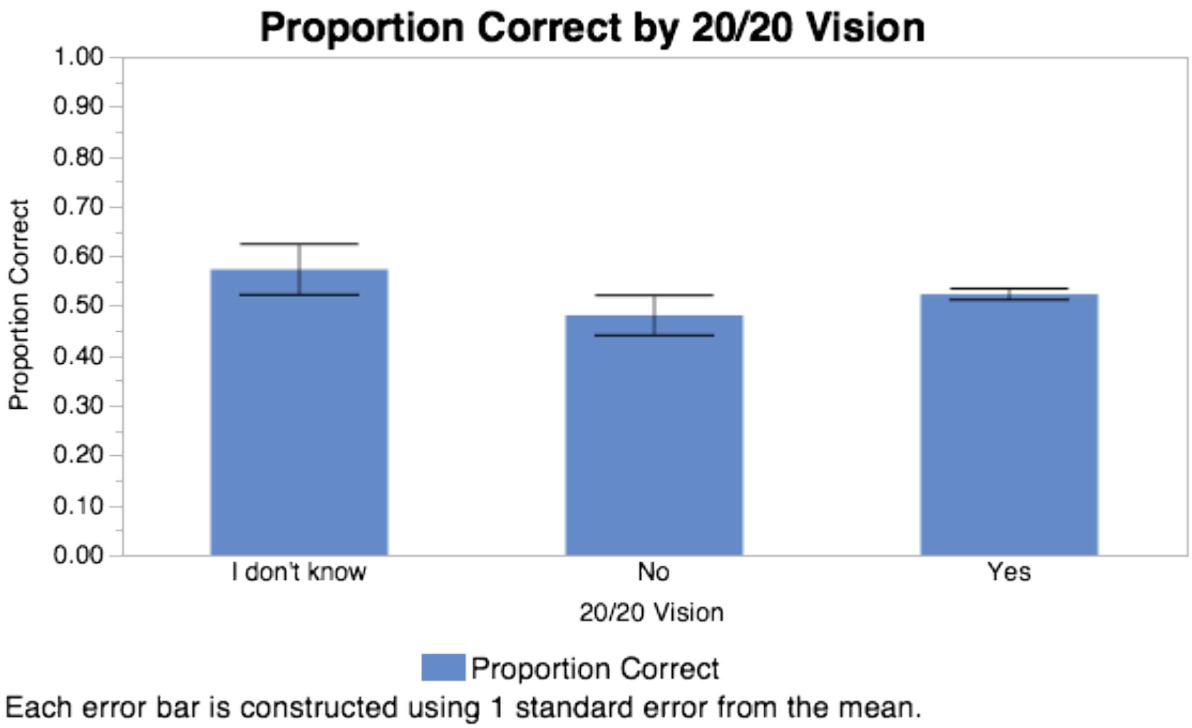


Figure 27. Bar Graph of Proportion Correct by 20/20 Vision.

No significance was found for 20/20 vision, as seen in Figure 27.

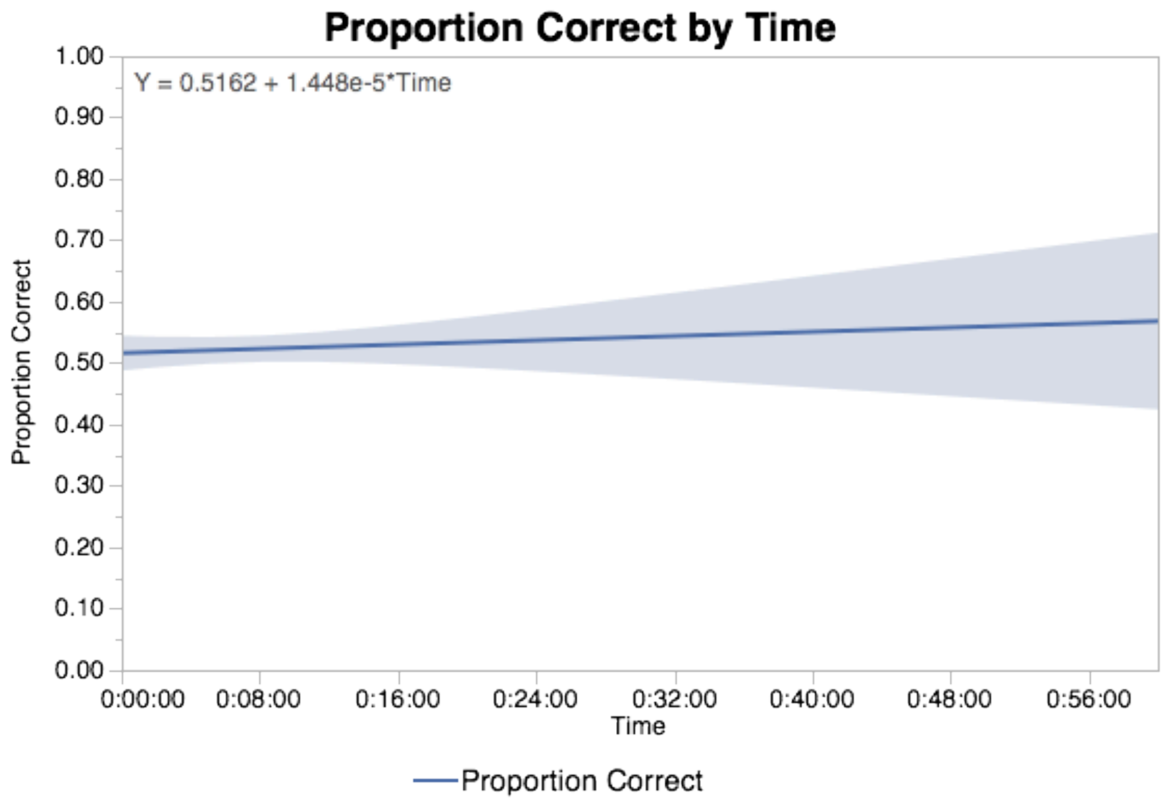


Figure 28. Proportion Correct by time to complete survey.

Figure 28 illustrates the trend between time taken on the survey and the proportion of correct responses. The blue area shaded around the trendline represents standard error of proportion correct. No significance was found for time taken on survey as a factor impacting performance. Most participants took under 24 minutes and those who took longer were not likely to perform better.

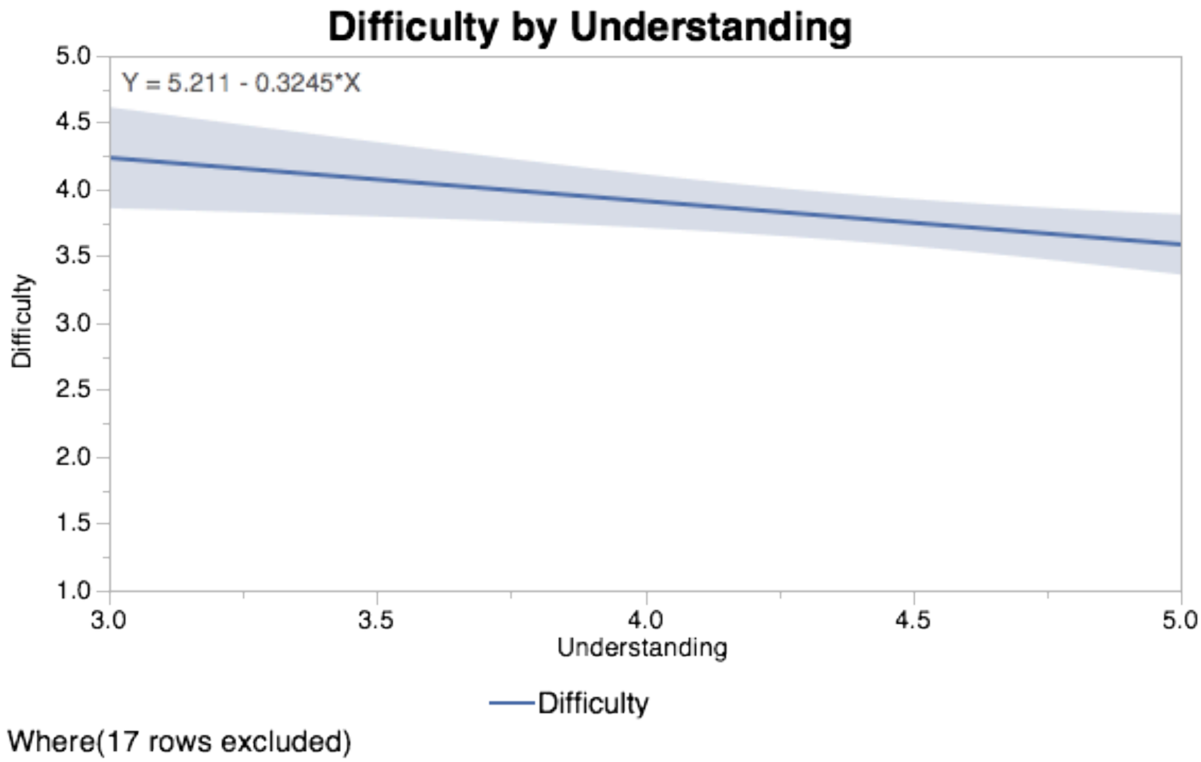


Figure 29. Difficulty vs. Understanding.

Figure 29 illustrates the trend between participant self-rated understanding of survey instructions and their rating of survey difficulty. The blue area shaded around the trend line represents standard error of difficulty. Participant rating of difficulty appeared to be only slightly inversely proportional to their rating of understanding.

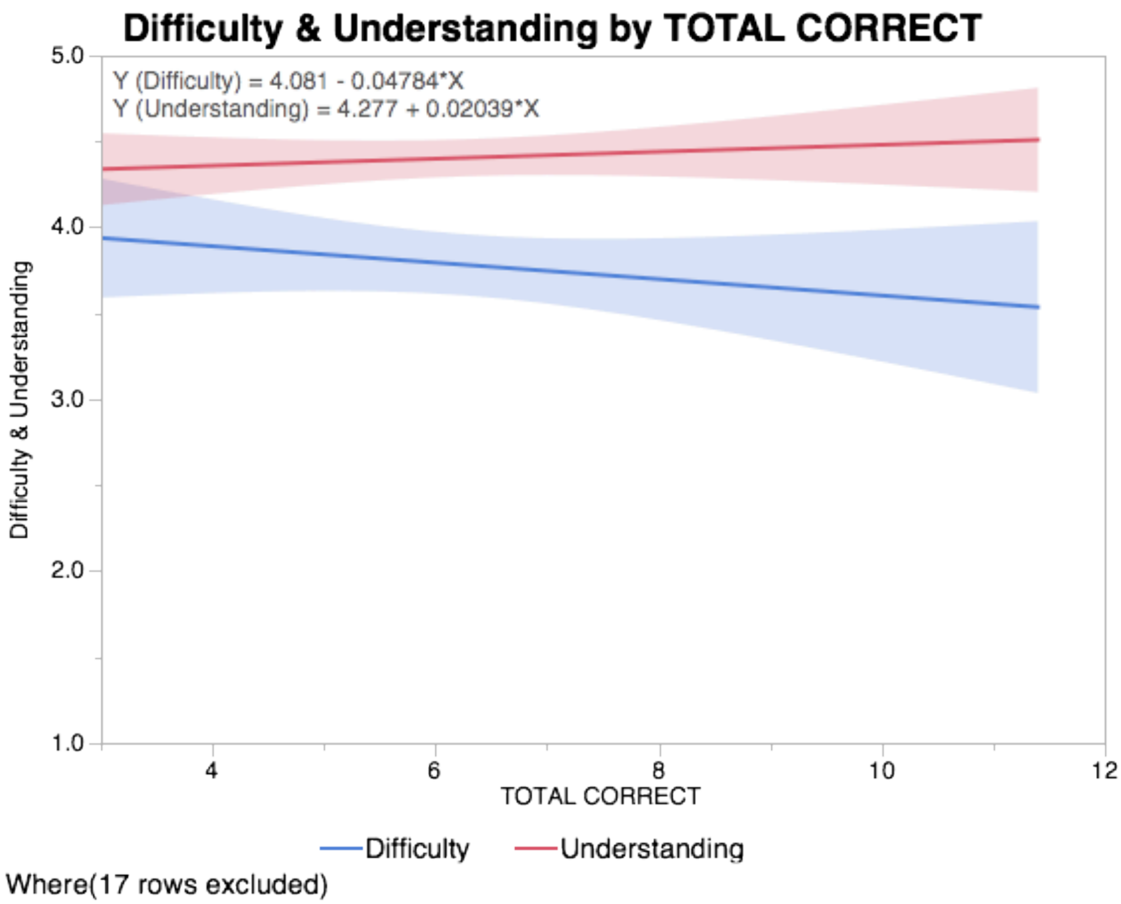


Figure 30. Comparing Total Correct Responses by Difficulty and Understanding.

It appeared in Figure 30 that as participant rating of understanding increased, their rating of difficulty decreased. However, neither self-rated difficulty nor understanding had a significant impact on whether a participant was able to perform better in detecting distortion.

### ***Observations from Primary Variables***

In addition to JMP, standard SAS was used for the last part of the analysis, determining the significance of the fixed effects, their interactions, and the random effects from blocking by person. Tables 16, 17, and 18 describe the Generalized Linear Mixed Model (GLMM).

## Results: ProcMixed.sas
## The GLIMMIX (GLMM) Procedure

**Table 16: Model Information**

| | |
|---|---|
| Data Set | WORK.FINALDATA |
| Response Variable | Selected |
| Response Distribution | Gaussian |
| Link Function | Identity |
| Variance Function | Default |
| Variance Matrix | Blocked |
| Estimation Technique | Restricted Maximum Likelihood |
| Degrees of Freedom Method | Containment |

**Table 17: Class Level Information**

| Class | Levels | Values |
|---|---|---|
| Label | 2 | Folgers Yogurt |
| Bottle Shape | 3 | Concave Conical Convex |
| Distortion | 4 | 25 50 100 12.5 |
| Person | 183 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 18 19 20 21 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 53 54 55 56 57 58 59 62 63 64 65 66 68 69 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 |

**Table 18: Covariance Parameter Estimates (Random Effects)**

| Cov Parm | Estimate | Standard Error |
|---|---|---|
| Person | 0.005525 | 0.002726 |
| Residual | 0.2366 | 0.0075 |

**Table 19: Type III Tests of Fixed Effects**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| Label | 1 | 1991 | 10.42 | 0.0013 |
| Bottle Shape | 2 | 1991 | 0.94 | 0.3917 |
| Bottle Shape*Label | 2 | 1991 | 5.24 | 0.0053 |
| Distortion | 3 | 1991 | 9.22 | <.0001 |
| Distortion*Label | 3 | 1991 | 0.93 | 0.4255 |
| Distortion*Bottle Shape | 6 | 1991 | 4.71 | <.0001 |
| Distortion*Label*Bottle Shape | 6 | 1991 | 1.38 | 0.2174 |

Table 19 shows that "Label" and Distortion were significant variables, and that there was an interaction effect between Bottle Shape and "Label" and between Bottle Shape and Distortion.

*Label*

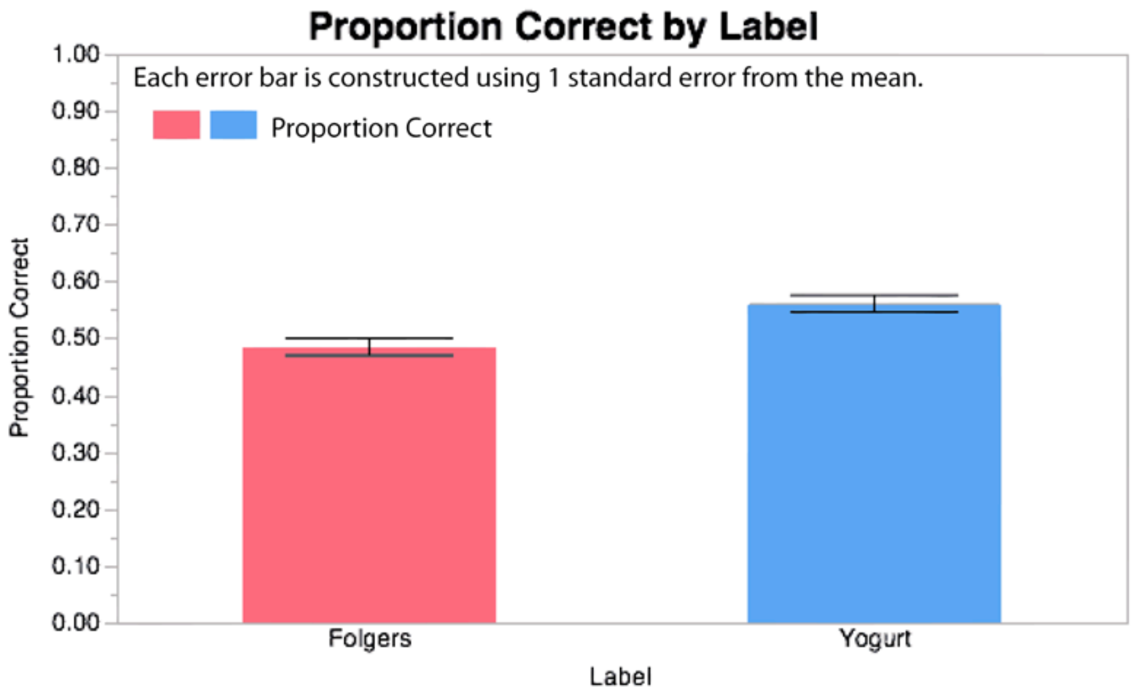


Figure 31. Bar Graph of Proportion Correct by Label.

**Table 20: Label LS-Means**

| Label | Estimate | Std Error | DF | t Value | Pr > \|t\| | Mean | Std Error Mean |
|---|---|---|---|---|---|---|---|
| Folgers | 0.4846 | 0.01657 | 1991 | 29.25 | <.0001 | 0.4846 | 0.01657 |
| Yogurt | 0.5604 | 0.01666 | 1991 | 33.65 | <.0001 | 0.5604 | 0.01666 |

**Table 21: Differences of Label LS-Means**

| Label 1 | Label 2 | Estimate | Std Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Folgers | Yogurt | -0.07584 | 0.02349 | 1991 | -3.23 | 0.0013 |

“Label” had a significant effect on detection of distortion, seen in Figure 31. The generic yogurt label had a better proportion of correct responses than the brand name coffee label. This could be due to label features that are more easily processed by the HVS, making distortion more obvious. For example, FDC recommends not using circles as they quickly lose their uniformity, and the generic yogurt label has an abstract circle of colorful dots. The HVS may recognize this pattern easily, as it can the triangle shown in Figure 4. It also may be that familiarity with a label is not as important as the features of the label itself. Table 20 gives the mean detection rate of distortion per “Label” and Table 21 indicates that these means are significantly different.

*Bottle Shape*

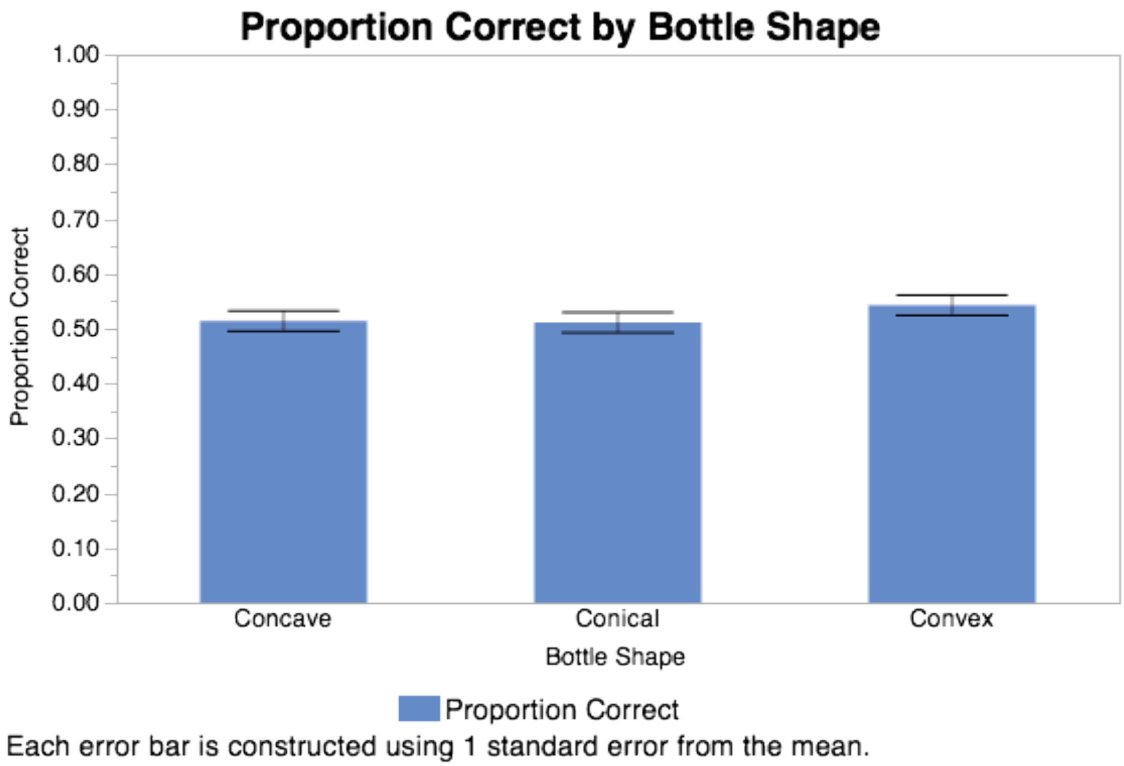


Figure 32. Bar Graph of Proportion Correct by Bottle Shape.

**Table 22: Bottle Shape LS-Means**

| Bottle Shape | Estimate | Std Error | DF | t Value | Pr > \|t\| | Mean | Std Error Mean |
|---|---|---|---|---|---|---|---|
| Concave | 0.5136 | 0.0188 | 1991 | 27.32 | <.0001 | 0.5136 | 0.0188 |
| Conical | 0.5113 | 0.0188 | 1991 | 27.2 | <.0001 | 0.5113 | 0.0188 |
| Convex | 0.5426 | 0.0188 | 1991 | 28.86 | <.0001 | 0.5426 | 0.0188 |

**Table 23: Differences of Bottle Shape LS-Means**

| Bottle Shape 1 | Bottle Shape 2 | Estimate | Std Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Concave | Conical | 0.002284 | 0.02543 | 1991 | 0.09 | 0.9284 |
| Concave | Convex | -0.02895 | 0.02543 | 1991 | -1.14 | 0.2550 |
| Conical | Convex | -0.03124 | 0.02543 | 1991 | -1.23 | 0.2194 |

Though Grahl, Greiner, & Walla (2012) found “evidence for gender-specific unbiased emotion differences related to different bottle shapes,” and gender did have an effect on detecting distortion in this final study, bottle shape did not have any effect outright (see Figure 32) nor an interaction effect with gender. Table 22 gives the mean detection rate of distortion per Bottle Shape and Table 23 indicates that these means are not significantly different.

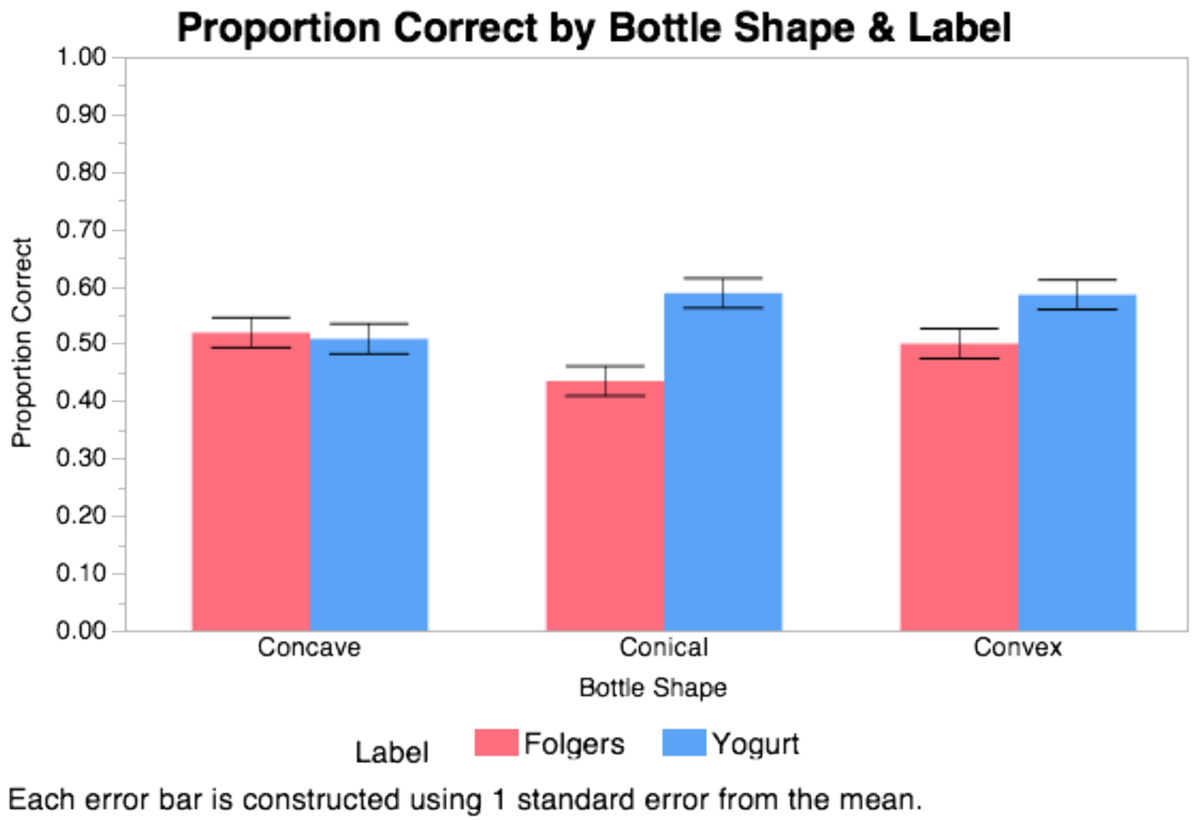


Figure 33. Bar Graph of Proportion Correct per Bottle Shape and Label

Although bottle shape did not have an effect outright, it did have an interaction effect with label, as seen in Figure 33. The effect was largest for the convex and conical bottle shapes, which promoted the identification of distortion when paired with the generic yogurt label. This may be attributed to features on the generic label being emphasized by the geometry of the convex and conical bottles. The brand name coffee label, though more familiar, perhaps did not have features that would promote detection of distortion when applied to the bottle geometries.

*Distortion*

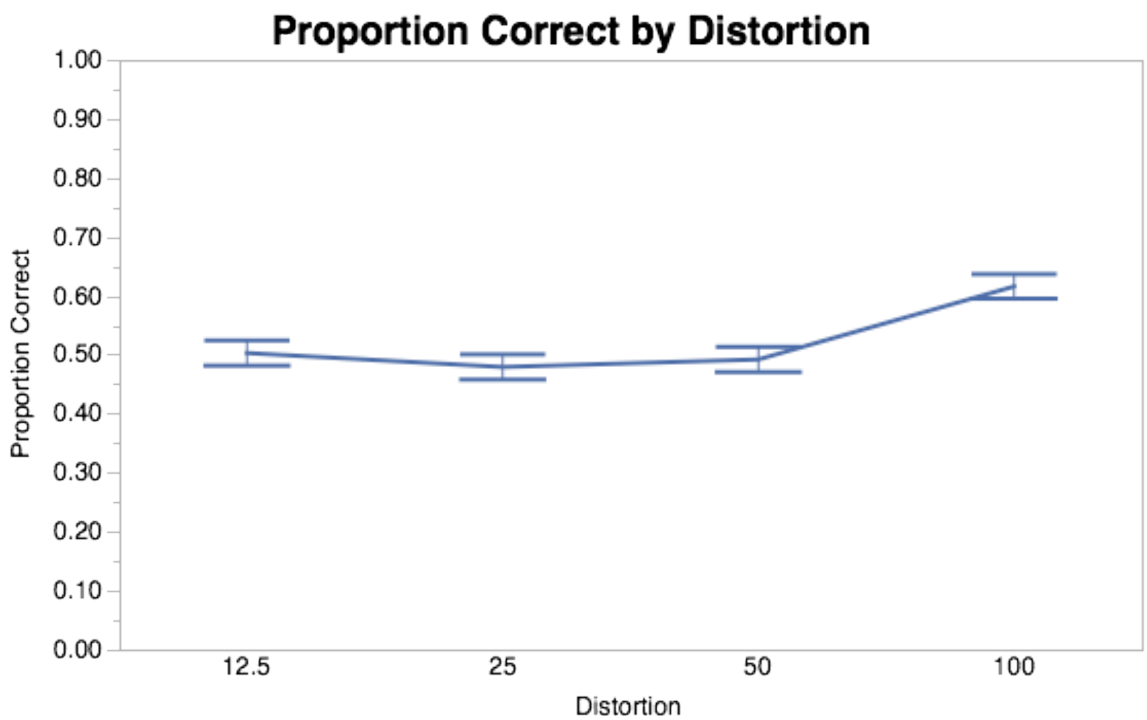


Figure 34. Line Graph of Proportion Correct per Distortion Percentage.

**Table 24: Distortion Percentage LS-Means**

| Distortion | Estimate | Std Error | DF | t Value | Pr > \|t\| | Mean | Std Error Mean |
|---|---|---|---|---|---|---|---|
| 12.5 | 0.5028 | 0.02148 | 1991 | 23.41 | <.0001 | 0.5028 | 0.02148 |
| 25 | 0.4792 | 0.02148 | 1991 | 22.31 | <.0001 | 0.4792 | 0.02148 |
| 50 | 0.4921 | 0.02148 | 1991 | 22.91 | <.0001 | 0.4921 | 0.02148 |
| 100 | 0.6160 | 0.02148 | 1991 | 28.68 | <.0001 | 0.6160 | 0.02148 |

**Table 25: Differences of Distortion Percentage LS-Means**

| Distortion 1 | Distortion 2 | Estimate | Std Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| 100 | 12.5 | 0.1132 | 0.02936 | 1991 | 3.85 | 0.0001 |
| 100 | 25 | -0.1368 | 0.02936 | 1991 | -4.66 | <.0001 |
| 100 | 50 | -0.1239 | 0.02936 | 1991 | -4.22 | <.0001 |
| 50 | 12.5 | -0.01073 | 0.02936 | 1991 | -0.37 | 0.7148 |
| 50 | 25 | -0.01286 | 0.02936 | 1991 | -0.44 | 0.6614 |
| 25 | 12.5 | -0.02359 | 0.02936 | 1991 | -0.8 | 0.4218 |

Figure 34 shows the percentage of graphic distortion and corresponding detection of distortion. Table 24 gives the mean detection rate of distortion per level of distortion and Table 25 indicates that the mean at 100% is significantly different than the means of all other levels of distortion. Up until 50%, distortion did not impact detection, but between 50 and 100% participants began to detect distortion. 100% distortion produced a mean correct response rate of 61.6%, which is less than what would be considered a threshold by classical methods. This implies that a strong Just-Noticeable-Difference (JND) threshold above no stimulus, or an absolute threshold, cannot be determined by classical threshold measures. Distortion still has a significant effect, however, once it is great enough. Consumers may be able to detect distortion on FBSS labels that have not been pre-distorted, but not to such an extent that it would prevent them from making a purchase decision. What's more, if any pre-distortion method is available, even a small application would be sufficient to mask any distasteful effects of distortion from the sleeve application process.

In conjunction with the significance of label familiarity, it may be true that a new, generic label can be introduced into a product market using FBSS and fare well enough without use of pre-distortion techniques, assuming label features attract attention as well as a brand name label but without drawing scrutiny to distortion. It would take a high level of distortion, the presence of certain unknown label features, and a certain choice of bottle shape for distortion to become evident. Perhaps distortion would be more perceivable on bottles with greater contours than those presented with this study, which would naturally distort graphics further. Without a threshold value, distortion can be only a minor concern when moving to FBSS labels on contoured bottle shapes.

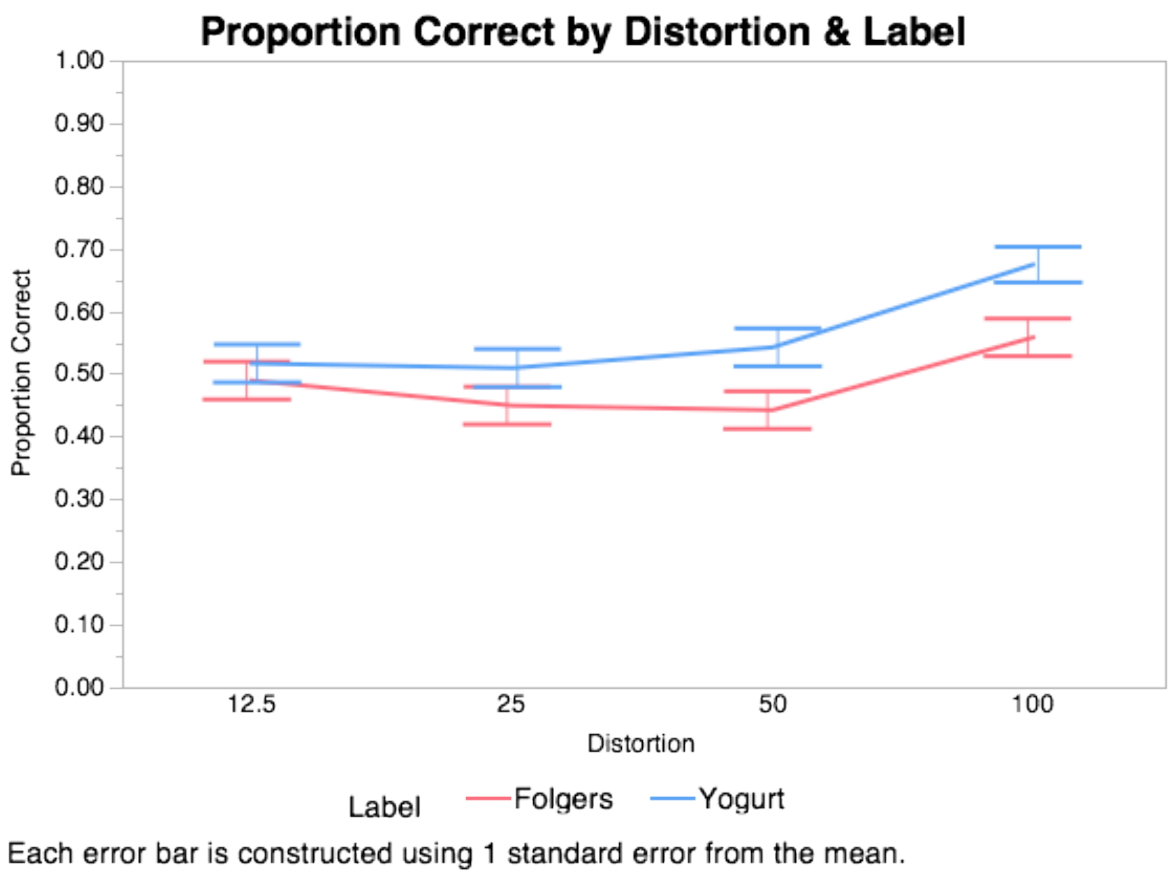


Figure 35. Line Graph of Proportion Correct per Distortion Percentage and Label.

There was not a significant interaction effect between percent distortion and label. However, in Figure 35 it may be seen that label alone did have an effect on detection of distortion, with the generic yogurt label having a mildly increased response rate.

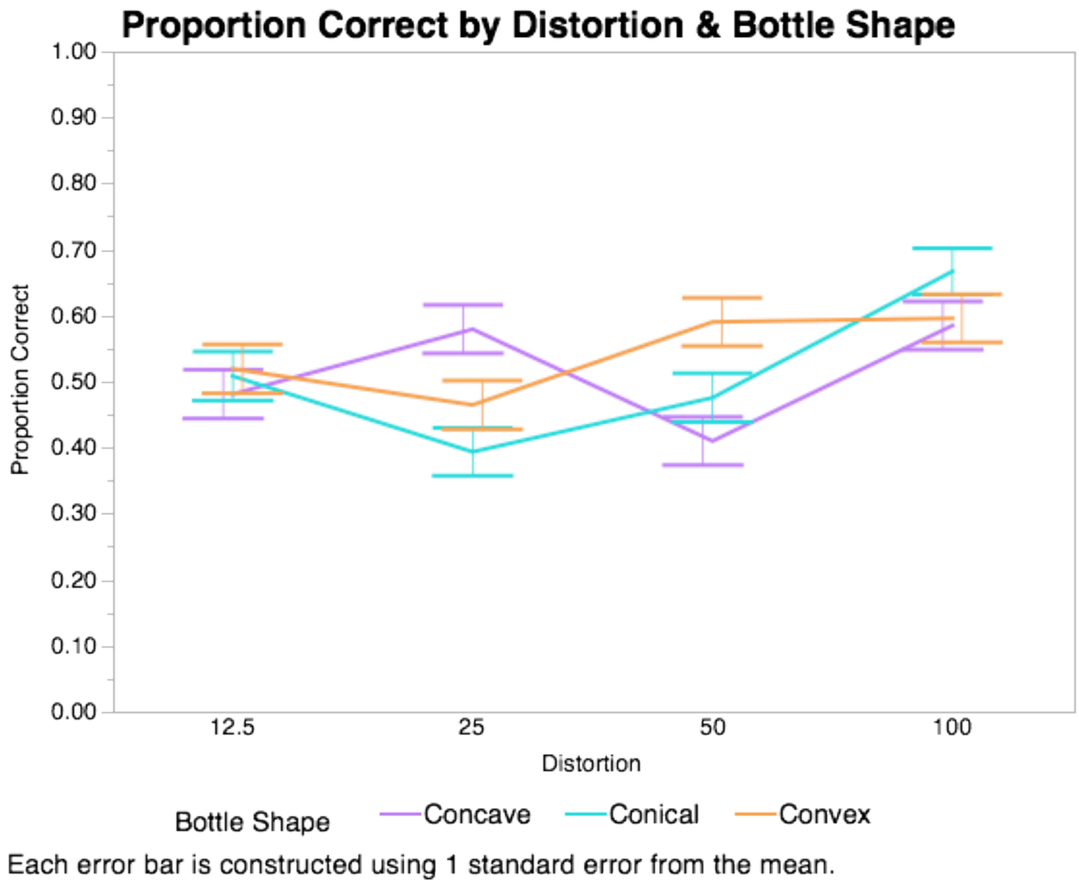


Figure 36. Line Graph of Proportion Correct per Distortion Percentage and Bottle Shape.

In Figure 36, there was a significant interaction effect between percent distortion and bottle shape. The concave bottle shape encouraged detection of distortion at 25% distortion, but discouraged it at 50% distortion. For all bottle shapes, 12.5% distortion had no better rate of detection than guessing, and 100% distortion had better rates of detection than guessing. Only between 25% and 50% does this anomaly present itself. This may be the inverted geometry of the concave bottle producing some masking effect with the HVS, or due to some other unknown cause. No errors in the assembling of the survey were found when investigated.

**Table 26: Placebo Response Location - Chi-Squared Test.**

| | Concave | Conical | Convex |
|---|---|---|---|
| Total Left | 110 | 110 | 103 |
| Total Right | 90 | 90 | 97 |
| Expected Each | 100 | 100 | 100 |
| P-value | 0.351691819 | | |

Table 26 shows that there was no position-response bias as determined by placebo questions, where neither of the two bottles presented had graphic distortion (full pre-distortion applied).

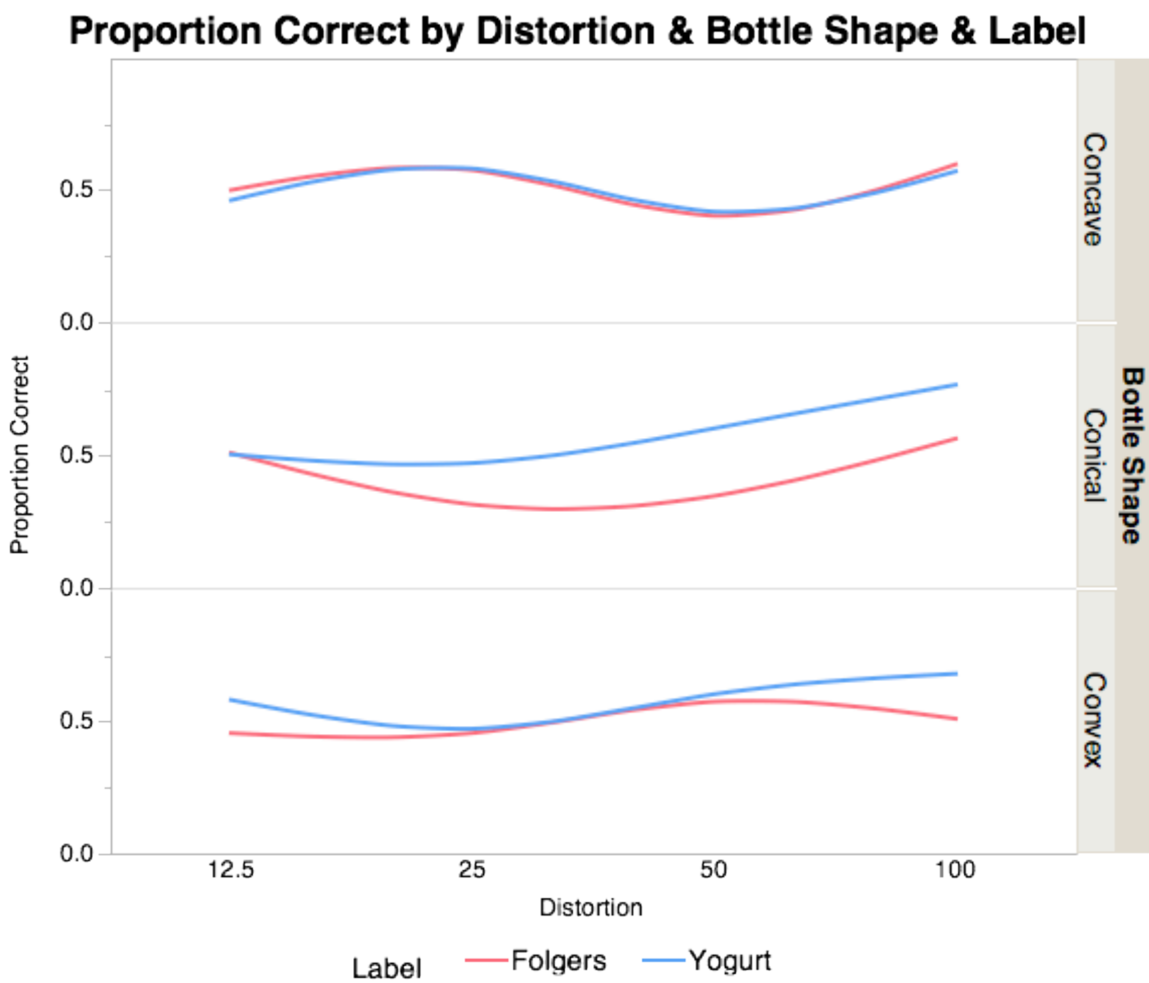


Figure 37. Line Graph of Proportion Correct per Distortion Percentage, Bottle Shape, and Label.

As seen in Figure 37, there was not a significant interaction effect between percent distortion, bottle shape, and label. However, here it may be seen that the interaction effect between percent distortion and bottle shape discussed above is very consistent between labels. This confirms that the effect stems either from the inverted geometry of the concave bottle, producing some masking effect with the HVS, or else from some other unknown cause.

# CHAPTER SIX

# CONCLUSIONS

[RQ1]: Is there a measurable threshold of acceptable distortion?

Fail to Reject $H_0$. Highest Response Rate of $0.616 < 0.75$

There is insufficient evidence, at 5% level of significance, to conclude that any means are greater than or equal 0.75, or a 75% detection rate, or to conclude that there is a measurable threshold of acceptable distortion.

$H_0$: $\mu_d$ <0.75, where d is some amount of distortion

$H_a$: $\mu_d$ ≥0.75, where d is some amount of distortion

No, under the simulated shopping conditions of this experiment there is not a measurable threshold of acceptable distortion, according to classical psychophysical threshold testing measures. A threshold would be need to be observed 75% of the time, the response rate halfway between 50% and 100% of the time. (Gescheider, 1997; ASTM E1808) At 61.6% response rate of detection for 100% distortion, the HVS does not detect any amount of distortion significantly often enough under the simulated shopping conditions of this experiment to determine a threshold for FBSS labels.

Reject $H_0$. P-value $<0.0001 < 0.05$

There is sufficient evidence, at 5% level of significance, to conclude that not all means are equal, and to conclude that consumers may detect distortion better at some levels.

$H_0$: $\mu_0=\mu_{12.5}=\mu_{25}=\mu_{50}=\mu_{100}$

$H_a$: $\mu_0\neq\mu_{12.5}\neq\mu_{25}\neq\mu_{50}\neq\mu_{100}$

However, when graphics on FBSS labels are distorted 100% (no pre-distortion applied), consumers may detect distortion significantly better than they otherwise would with lower levels of distortion. This is not a threshold amount of distortion, but simply a measurable amount that is significant in effect. This may indicate that 100% distortion is just short of significance as a threshold, but significant enough to have an effect of some concern. This finding may allow less rigid brand regulations and increased application of graphics to high distortion areas of bottles. High consumer tolerance of distortion may allow reduced testing costs and time to market for FBSS.

P-value $<0.0001 < 0.05$

Distortion also has a statistically significant interaction effect with bottle shape, apparent most on conical and convex bottles.

[RQ2]: Does consumer preference for non-distorted graphics affect consumer acceptance of distortion?

Fail to Reject $H_0$. P-value $0.4990 > 0.05$

There is insufficient evidence, at 5% level of significance, to conclude that the means are different, or to conclude that consumer preference for non-distorted graphics affects consumer acceptance of distortion.

$H_0$: $\mu_{Yes}=\mu_{No}$

$H_a$: $\mu_{Yes}\neq\mu_{No}$

12% of participants said that they do prefer non-distorted graphics. 22 out of 183 participants said distortion would impact their purchase decisions, and 161 out of 183 said it would not. Figure 16 illustrates that even if they self-report that distortion would impact their purchase decisions, their sensitivity to distortion remained the same. This is perhaps because they physically cannot detect distortion, are not attentive enough to detect distortion on FBSS, or find their preference for non-distorted graphics to be less than they believed. Consumers who thought distorted graphics would prevent them from making a purchase may be put at ease knowing that graphics do not directly indicate damaged products. FBSS may be more widely used in the future with less uncertainty about whether distorted graphics could negatively affect consumer purchase likelihood.

[RQ3]: Does bottle shape affect consumer acceptance of distortion?

Fail to Reject $H_0$. P-value $0.3917 > 0.05$

There is insufficient evidence, at 5% level of significance, to conclude that any of the means are different, or to conclude that bottle shape affects consumer acceptance of distortion.

$H_0$: $\mu_{Convex}=\mu_{Concave}=\mu_{Conical}$

$H_a$: $\mu_{Convex}\neq\mu_{Concave}\neq\mu_{Conical}$

Bottle shape alone does not seem to affect consumer acceptance of distortion, but bottle shape may have an effect when it interacts with labels and with percentages of distortion.

[RQ4]: Does familiarity with graphics affect consumer acceptance of distortion?

Reject $H_0$. P-value $0.0013 < 0.05$

There is sufficient evidence, at 5% level of significance, to conclude that the means are different, and to conclude that familiarity with graphics does affect consumer acceptance of distortion.

$H_0$: $\mu_{BrandName}=\mu_{Generic}$

$H_a$: $\mu_{BrandName}\neq\mu_{Generic}$

Yes, in fact familiarity may cause a masking effect or blindness to distortion. Consumers may see brand name products they know and pay no further attention to a label or if it is distorted. Also, the masking effect may depend on certain features that make up the label. This effect may be strongest for labels that have more features, edges and contrast, or labels without excessive visual noise but with details that are very specific, like the profile of a rabbit or a face.

P-value $<0.0053 < 0.05$

Label also has a statistically significant interaction effect with bottle shape, apparent most on conical and convex bottles.

### *Limitations*

Because this final study used only 2D digital simulations of tangible 3D objects, its conclusions may be limited in scope. This did, however, allow for more controlled samples and for a broader test subject base than would have been accessible otherwise. Results may only apply to the bottle types specified here: concave, convex, and conical shapes. Asymmetric, ergonomic, and bottles generally proportioned differently may also have a different effect with distortion. Familiarity with label graphics was determined to impact detection of distortion, though graphics on labels can admittedly be a diverse and subjective craft, and it cannot be assumed that these conclusions are universal.

### *Recommendations*

The only demographic variable that had a significant fixed effect was gender, and the distribution of the histogram for all participants was bimodal when sorted by gender. Men identified distortion significantly better, possibly because they tend to be more visually oriented, more image-driven in their preferences for specific designs, and more concerned with overall product structure, whereas women pay more attention to details, textures and natural-looking forms. Men generally view design simplicity in image-related terms, while women prefer it in practical terms. Rather than being more perceptive, it may be that men were more competitive and had more personal investment in getting as many correct as possible, while women were simply finishing the survey for compensation. This difference in gender should certainly be investigated further and included in any future study.

“Label” had a significant effect on detection of distortion. The generic yogurt label had a better proportion of correct responses than the brand name coffee label. This could be due to label features that are more easily processed by the HVS, making distortion more obvious. It may be that familiarity with a label is not as important as the features of the label itself. Both labels were modeled after actual labels of products that are typical with FBSS. Further research is needed on label familiarity and on the elements of labels themselves. Research could expand to include new products that are not yet typical with FBSS.

Although bottle shape did not have an effect outright, it did have an interaction effect with “label.” The effect was largest for the convex and conical bottle shapes, which promoted the identification of distortion when paired with the generic yogurt label. This relationship between unfamiliar, generic labels and bottle shape is worth pursuing further.

There was a significant interaction effect between percent distortion and bottle shape. The concave bottle shape encouraged detection of distortion at 25% distortion, but discouraged it at 50% distortion. Due to this peculiar outcome, similar bottles with more exaggerated curvatures and new bottle shapes should be included in future research. Bottles with greater contours than those presented in this study would naturally distort graphics further, making distortion more likely to be perceived.

Other possible investigations include measuring the minimum amount of time required to identify distortion in an image (Thorpe, Fize, & Marlot, 1996), or using eye tracking technology to measure where and how people determine distortion. Moving to physical FBSS prototypes in a simulated shopping environment is the next logical advancement of this body of research. Further study of shrink-wrap technologies and control capabilities, especially the more common steam-based shrinking process, would aid further experiments with this direction and application.

Companies might consider using less of their logos on bottles that utilize FBSS, as proposed by Ewam Yap with Big Brand Theory: Experimental Packaging Design (2013). “Each brand’s identity could be meticulously and uniquely cropped out of the packaging as much as possible, maintaining its integrity and comprehension while at the same time enhancing the aesthetic value.” With less actual logo, the FBSS labeling process would be simpler while maintaining consumer brand recognition.

The classical method of measurement did not yield a threshold, but the Generalized Linear Mixed Model (GLMM) did find that distortion could be a significant factor perceived by consumers. Further comparison between methods of threshold testing should be done, especially when it relates to graphic distortion on Full Body Shrink Sleeves (FBSS). Quality assurance sensory procedures aim to ensure consumer preferences are satisfied, but if a threshold is beyond consumer perception or variance in quality control, then consumer rejection from said sensory quality should not be a concern.

# APPENDICES

## Appendix A:

## Amazon's Mechanical Turk

The 21st century saw the first use of online surveys for data collection purposes (Birnbaum, 2000). Amazon's Mechanical Turk (MTurk) was introduced in 2005 as a "marketplace for work that requires human intelligence," pairing together "requestors" and "workers" for short-term tasks. (www.mturk.com) Peer-reviewed articles on the strengths and weaknesses of the tool quickly followed the first appearance of MTurk data the research literature. (Rouse, 2015) Institutional Review Boards (IRBs) have naturally raised specific questions about use of MTurk in academic research, though it has all the major elements required to conduct research from start to finish: a streamlined process of study design, a large participant pool, facilitated participant recruitment, data collection, and integrated participant compensation system. (Buhrmester et al, 2011) Additionally, MTurk offers numerous benefits such as the lack of potentially biasing interactions with the experimenter, access to a distinctly different population than traditional college samples, built-in anonymity, and availability of cross-national data. (Rouse, 2015)

In what appears to be the first psychological research publication using MTurk, Eriksson and Simpson (2010) found that both American and Indian women expressed stronger negative emotional reactions to the prospect of losing money in a lottery, relative to American and Indian men. This replicated a gender difference in financial risk-aversion in a general online setting. Not long after, Alter, Oppenheimer, and Zemla (2010) showed that both MTurk workers and traditional lab participants exhibit the same cognitive error of tending to overestimate one's ability to explain natural and mechanical processes. In a replication of classic decision-making research tasks, Paolacci, Chandler, and Ipeirotis (2010) also showed that rates of judgment errors did not differ significantly for MTurk workers in comparison to more conventional samples.

### *Participation, Compensation & Length*

The ease and relatively low cost of gaining the data can be particularly beneficial to collaborating student and faculty researchers. Data reported by Rouse (2015) for a study included 400 responses, was collected in 13 hours and 13 minutes, and cost $88. Chris Callison-Burch, PhD, an associate research professor in the computer science department at Johns Hopkins University in Baltimore, has used MTurk for his research. (Reviewing Research, 2010)

Amazon charges a 10% commission, and the compensation amount awarded in MTurk is typically small, as low as $0.01 and rarely exceeding $1, for 5-10 minute Human Intelligence Tasks (HITs). Translated into an hourly wage, the typical worker is willing to work for about $1.40 an hour (Horton & Chilton, 2010). To put costs in perspective, a pay rate on MTurk of $.50 for a 5-min survey (an effective hourly rate of $6.00) is still associated with a per-respondent cost of $.55 (including Amazon.com's 10% surcharge) or $.11 per survey minute. By contrast, per subject costs for typical undergraduate samples are about $5–10, for nonstudent campus samples about $30 (Kam, Wilking, and Zechmeister 2007), and for temporary agency subjects between $15 and $20. (Berinsky et al, 2012)

Realistic compensation rates do not affect data quality, and the data obtained are at least as reliable as those obtained using traditional methods. However, participation rate is a function of HIT features, payment amount and survey length, a finding consistent with previous research on non-survey tasks (Mason & Watts, 2009; Buhrmester et al, 2011). When offering just 2 cents for a 30-minute task, 25 participants can be recruited in about 5 hours of posting time. By increasing the compensation to 50 cents, the same number of participants may be recruited in less than 2 hours of posting time. (Buhrmester et al, 2011) Generally, recruiting is fast and it took only three weeks for Paolacci et al to collect 1000 subjects. Each respondent was paid $0.10 and took 3 minutes on average to complete the survey, an hourly average wage of $1.66, which is superior to the median reservation (minimum acceptable) wage of $1.38/hour (Horton & Chilton, 2010).

Although no MTurk worker is obligated to complete a HIT, the ethical principle of Justice compels researchers to provide a fair reward. (American Psychological Association, 2010) IRBs usually require that a person be paid regardless of whether he or she withdraws from a study. A preemptive qualification could have participants answer questions unrelated to the research demonstrating their attention, and dismissing them from the study if they don't pass the screening test. (Reviewing Research, 2010). Workers are willing to complete simple tasks for virtually no compensation, suggesting that they are not driven primarily by financial incentives, but by internal motivations such as challenges and enjoyment. (Buhrmester et al, 2011)

### ***Demographics & Motivation***

When compared to the practical alternatives, the MTurk respondent pool has attractive characteristics, even apart from issues of cost. MTurk boasts a large, diverse workforce consisting of over 100,000 users from over 100 countries who complete tens of thousands of tasks daily (Pontin, 2007). They are also not currently an excessively overused pool. Heavy reliance on American college samples, generally small sectors of humanity, has been a longstanding gripe, especially in the field of psychology (Buhrmester et al, 2011). Although not without it's own faults, data collected via the Internet can reduce the biases of traditional samples. MTurk participants are more demographically diverse than standard Internet samples and significantly more diverse than student and convenience samples. (Gosling et al, 2004; Paolacci et al, 2010)

70–80% of workers were from the United States and were relatively representative of the population of U.S. Internet users. Recently, however, the population dynamics on MTurk have changed significantly, with a greater proportion of Indian subjects in recent experiments, suggesting the need for a fresh survey of the workers. (Eriksson & Simpson, 2010) Amazon allows participation in a given HIT to be restricted to workers from a specific country, allowing researchers to maintain a homogeneous population despite growing heterogeneity. (Paolacci et al, 2010) On race, MTurk's characteristics are mixed: It is similar to the Current Population Survey (CPS) on percent white, but underrepresents blacks and Hispanics. (Berinsky et al, 2012)

Across U.S.-based MTurk workers, there are significantly more females (64.85%) than males (35.15%). (Paolacci et al, 2010) Compared to the American National Election

2008–09 Panel Study (ANESP), a high-quality Internet survey conducted by the firm Knowledge Networks, MTurk is only slightly more female (64 versus 58%), and slightly less educated (14.9 versus 16.2 years). (Berinsky et al, 2012) Though it is self-reported, both MTurk and ANESP somewhat underrepresent low education respondents. This may partially be explained by the younger age of MTurk users, but may also reflect higher education levels among early adopters of technology. On age, the MTurk sample is notably younger. In the study by Paolacci et al. (2010) workers were 36 years old on average, with a minimum of 18, maximum of 81, and median of 33, and thus slightly younger then both the U.S. population as a whole and the population of Internet users.

Despite being more educated, both mean and median income for MTurk are lower than ANESP, CPS and national averages. (U.S. Census, 2007) For example, while 45% of the U.S. Internet population earns below $60K/year, the corresponding percentage across U.S.-based MTurk workers is 66.7%. (Paolacci et al, 2010) Only 13.8% of the U.S.-based workers reported that MTurk was their primary source of income. 61.4% reported that earning additional money was an important driver for joining MTurk, with Entertainment at 40.7% and “killing time” at 32.3%. In fact, 69.6% of the U.S.-based workers reported that they consider MTurk a fruitful way to spend free time. Most workers spend a day or less per week working on MTurk, and tend to complete 20–100 HITs during this time.

Relative to ANESP and CPS, MTurk subjects are more likely to: have never married (51%), rent rather than own their home (53%), and report no religious affiliation (42%). The MTurk sample is broadly similar to the other samples in region of residence, with perhaps a slightly larger prevalence of those living in the Northeast. MTurk respondents are slightly more Democratic in their partisan identification than are ANESP respondents and are substantially more liberal in their ideology. (Berinsky et al, 2012)

All told, these comparisons reinforce the conclusion that the MTurk sample does not perfectly match the demographic and attitudinal characteristics of the U.S. population, which may limit their suitability for some research topics, but does not present a wildly distorted view of the U.S. population either. (Berinsky et al, 2012) MTurk samples will often be more diverse than convenience samples and will always be more diverse than student samples. MTurk provides an important way to overcome the barrier to conducting research raised by subject recruitment costs and difficulties by providing easy, inexpensive access to nonstudent adult subjects. (Berinsky et al, 2012)

***Anonymity***

The entire interaction between investigator and subject is online and anonymous, reducing concerns about how to safely store responses to sensitive questions, and IRBs are more likely to treat studies in MTurk as exempt from reviews. (Paolacci et al, 2010) While a researcher may not know the name of the person assigned to a particular number, if the subject participates in many studies, a fairly detailed personal profile could still be derived from the answers. Amazon officials did not respond to a request from IRB Advisor about how the company protects data. (Reviewing Research, 2010) So, IRBs concerned about this possibility can require investigators to delete the worker's official Amazon ID number and assign them a different ID that is unique to a specific study once the data is collected. This could be a standard practice in IRB proposals.

Also because of anonymity, it is impossible to obtain a signed consent form. If the IRB permits it, a requester can include consent in the request, explicitly stating that the task is research and that acceptance of it constitutes giving consent. A more in-depth consent statement may be part of a screening tool, requiring participants to read the consent statement and click a button saying, "I consent" before being able to participate in the study itself. (Reviewing Research, 2010)

***Data Quality***

There is little evidence to suggest that data collected online is necessarily of poorer quality than data collected from subject pools (Krantz & Dalal, 2000; Gosling et al., 2004). The quality of data provided by MTurk meets or exceededs the psychometric standards associated with published research. (Buhrmester et al, 2011) Concerns about heterogeneous treatment effects, subject attentiveness, and the prevalence of habitual survey takers are not large problems in practice. Given their incentives, MTurk respondents may generally pay greater attention to experimental instruments and survey questions than do other subjects. (Berinsky et al, 2012)

In practice, multiple responses are rare in web based experiments and are even less of a problem for MTurk because each worker ID must correspond to a unique credit card number. (Krantz & Dalal, 2000; Gosling et al., 2004, Berinsky et al, 2012) Thanks to the workers' unique ID, researchers can identify workers who already took part to previous versions of an experiment, and exclude them accordingly. (Paolacci et al, 2010)

Non-response error is the number of people who accessed the study but did not complete it entirely. Subjects recruited from online discussion forums are significantly less likely to complete the survey than MTurk subjects (66.7% and 91.6% respectively) This suggests that MTurk strongly diminishes the potential for non-response error in online research. (Paolacci et al, 2010) When using external sites, such as SurveyMonkey, it is important to remember that eventually the experimenter will need to verify workers' claimed participation. This can be done by providing a code at the end of the survey to be retrieved and entered back on the MTurk HIT page.

External validity is "an assessment of whether the causal estimates deduced from experimental research would persist in other settings and with other samples. Internal validity pertains to the question of whether causal estimates appropriately reflect the effects of the experimental manipulation among the participants in the original setting." (Berinsky et al, 2012) "Workers in MTurk exhibit the classic heuristics and biases and pay attention to directions at least as much as subjects from traditional sources. Furthermore, MTurk offers many practical advantages that reduce costs and make recruitment easier, while also reducing threats to internal validity." (Paolacci et al, 2010)

MTurk subjects appear to respond to experimental stimuli in a manner consistent with prior research. When experimental studies previously conducted using convenience and nationally representative samples were replicated, MTurk and original samples had similar results. (Berinsky et al, 2012) Overall, these studies confirm that MTurk is a reliable source of experimental data in judgment and decision-making. (Paolacci et al, 2010) Hundreds of studies are being conducted using this research method. (Rouse, 2015)

Appendix B:

Screen Size: Preference & Performance

***Preference***

When subjects watch the larger television screens they report more positive emotional responses to attractive news anchors on the screen and to viewing environment (Grabe et al, 1999) When subjects watch a dramatic program, they prefer larger television displays. (Ohtani & Mitshuhashi,1973) The research on screen size generally suggests that large screens increase the intensity of viewer responses, but some experiences may become too intense and even unpleasant on large screens. When viewing a horse race scene that features fast motion, larger displays cause dizziness and fatigue. (Ohtani & Mitshuhashi,1973) Three studies provide no support for the idea that screen size can enhance the enjoyment of a viewing experience; Kim (1996) Detenber & Reeves (1996) and Lombard et al (1997). Sense of distance, lack of realism, and decreased attention and memory are associated with small screen size viewing conditions. On a social level, it seems that small screens might be preferable in personal settings whereas large screens serve public or small group viewing experiences better. (Grabe et al, 1999)

***Viewer Disability***

A study was conducted comparing the preference of 3 screen sizes between disabled and nondisabled persons in a word processing task. (Li et al, 2010) Most disabled and nondisabled participants believed a larger screen made it easier to locate characters and spell, while smaller GPS and especially cell phone screens were very difficult to read and catch the stimuli of target characters. Participants experienced dry eyes and fatigue with a smaller screen more quickly than with larger screens. A bigger screen size could generally lead to better performance, though screen size does have a limit for participants with motor disabilities due to their eye span or visual field. However, a larger screen size will elicit more unnecessary movement for all people. Results supported the hypothesis that different screen sizes could be used with satisfactory performance for nondisabled participants. (Li et al, 2010)

***Quality & Conditions***

Some researchers have found that larger screens make people choose a greater viewing distance (Duncanson & Williams, 1973; Lombard, 1995; Lund, 1993; Nathan et al, 1985). Others have found the ratio of preferred distance to image size to be constant regardless of image size (Jesty, 1958; Westerink & Roufs, 1989). The work of Lund (1993) may be the most authoritative: in five experiments he showed subjects images from 11 to 123 inches and found the ratio of viewing distance to image height declined from 7.4 to 3.1. Preferred viewing distance may be related to subjective evaluations of image quality (i.e. - we sit closer to images we perceive to be of better quality).

When resolution is held constant, variations in screen size do not impact the amount of information displayed on a screen but make that information appear smaller (on a smaller monitor) or larger (on a larger monitor). Although the amount of information remains constant, test takers might find one easier to read than another.

(Bridgeman et al, 2003) No significant effects on math or verbal scores were found in a study with larger high-resolution displays. (Bridgeman et al, 2003)

Effects could be different in a high-stakes test. Frustrations with scrolling might reduce motivation and effort in a low-stakes test but have less of an impact in a higher-stakes test. On the other hand, difficulties with scrolling could raise anxiety levels and therefore have more of an effect on a high-stakes test. (Bridgeman et al, 2003) Because scrolling to search for answers takes time, larger effects of screen resolution might be found in a rapid test. The math test, which required virtually no scrolling, should be less susceptible to effects from stricter time limits. Nevertheless, with less time per item, screen display effects might have emerged. Studies by Bridgeman et al. (2003) were relatively short, and factors that are merely minor irritants on a short test might have more of an impact on longer tests.

***Memory***

Detenber and Reeves (1996) found that subjects who read from a 15 inch computer screen learned the material more quickly than those who read from a 12-inch screen. Yet, there were no differences in required cognitive effort or long-term retention of the material. They also found that subjects who watched images on a large screen remembered more of the images immediately after viewing than subjects who watched a small screen. However, screen size did not influence delayed memory. Kim (1996) found that a large image lead to greater memory for facts presented and greater recognition memory of images in a 15-minute infomercial.

Appendix C:

Input Code for Statistical Analysis in SAS

```
PROC IMPORT OUT= FinalData DATAFILE= FINALDAT
        DBMS=xlsx REPLACE;
    SHEET="sheet1";
    GETNAMES=YES;
RUN;

data FinalData;
     input Product Bottle Distort Person Selected @@;

          Proc Glimmix data = FinalData;
                    class Product Bottle Distort Person;
                    model Selected = Product Bottle Distort Distort*Bottle
                    Product*Bottle Product*Distort Product*Bottle*Distort;
                    random Person;
          LSmeans Product Bottle Distort/ilink;
          LSmeans Product Bottle Distort/diff;
          run;
quit;


*Some variable names were command function names and were changed for coding in
SAS
*Product = label
*Bottle = bottle shape
*Distort = distortion
*Person = participant

*  Basic GLM code
 *  proc glm data=FinalData;
 *    class Product Bottle Distort Person;
 *     model Selected=Product Bottle Distort Distort*Bottle Person;
 *     random Person / test;
 *  run;
*quit;
```

## Appendix D:

## Growth of Polyolefins

Polyolefins are the largest group of thermoplastics, produced from chemicals derived by processing coal, crude oil and natural gas. Olefin means "oil-forming" and these compounds are also known as alkenes. Polyolefins are built with unsaturated or reactive monomers that consist of only hydrogen and carbon atoms. These monomers also have at least one pair of double-bonded carbon atoms. These molecules are simple, but when repeated and combined at length in various ways, they produce materials with unique properties. The two most important types of these materials are polyethylene (PE) and polypropylene (PP).

Polyolefins are used as stand-alone plastic components and often in composite structures. Generally, a polyolefin material consists of polymer chains throughout its mass. These polymers can change depending upon how they are produced and especially upon temperature. If the polymer chains are amorphous, like LDPE in Figure 38, they have many randomly entangled molecule branches that form a flexible, transparent material. If polymer chains are non-branched and crystalline, like LLDPE or HDPE, they tend to organize closely around spherulite crystals forming a dense, rigid, and opaque or translucent structure. At high temperatures, specifically beyond the melting point ($T_m$), these structures are not maintained and the material exhibits amorphous appearance and behavior.

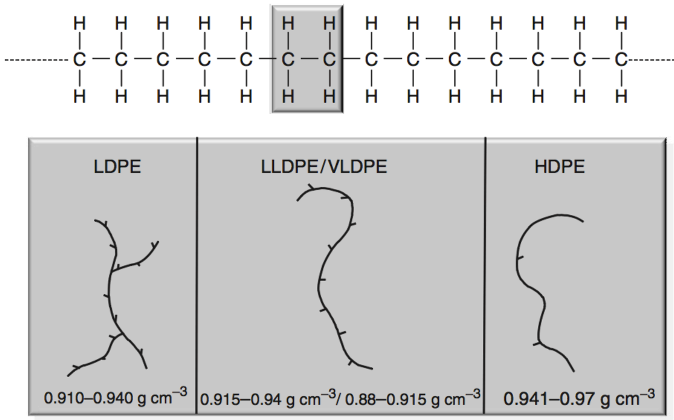


Figure 38. Polymer Chains in Polyethylene (PE). (Soares & McKenna, 2012).

Imperial Chemical Industries first commercialized polyolefins, but diverse properties allow their uses to include adhesives, appliances, automotive parts, cable jacketing, crates/containers, textiles, toys, film, food packaging, insulation, plastic bags, medical applications like prosthetics, textiles, and numerous other goods. It is now increasingly substituted for other materials in previous applications.

One implementation of interest is shrink-wrap film, where a polymeric plastic is useful for its ability to return to smaller dimensions when exposed to heat. Herbert Nagel and Lawrence Roy Hatt with DuPont (E.I. du Pont de Nemours and Company) based out of Willmington, Delaware, invented shrink-wrap film. They applied for a patent in 1955 and by March 3, 1959, Patent 2,876,067 made it official (Heat-shrinkable film, 2013). Since it's creation, production of shrink film has expanded to include multi-substrate films, which combine advantageous qualities, and novel materials in an effort to reduce disposal waste.

Due to industrialization in emerging markets, expanding global communications, and increased trade, the market for polyolefins is changing rapidly. New applications, especially where there are cost and performance advantages, will drive demand upward. Production has seen a lot of investment and growth (Figure 39) in places like the Middle East and especially Asia Pacific where there are great supplies of affordable raw materials and potential for high demand. This trend is expected through 2014. On the other hand, Europe's landscape is shifting to higher value and performance products while consolidating properties and operations, a process sometimes called "rationalization." Cheap feedstock from shale gas in North America is stimulating the production and market for PE, making it a globally competitive export (Tonzani, 2012; America's Bounty: Gas Works, 2012). However, the American and European markets are viewed as mature and are not expected to contribute much growth over the next decade. Profits for the industry are expected to climb until around 2015, at a cyclical high not reached since 2006. Global demand for olefins and polyolefins was around 205 and 116 million tons, respectively, in 2010 and is expected to grow by 4% annually until 2017, just after the peak demand (Plastemart, 2011). Asia Pacific accounts for nearly 40% of demand worldwide, with an expanding manufacturing sector generating exports and some domestic products. This region will continue to import large polymer quantities for the foreseeable future, with China leading much of the industry growth. This is partly due to local coal reserves and new refinery investments. India, Singapore, Malaysia and Thailand are expanding rapidly and there are signs of potential in Indonesia and Vietnam, where there are large populations and potential markets. Long-term progresses will be sustained by advanced feedstocks of natural gas liquids like ethane.

| Year | LDPE | LLDPE | HDPE | PP | Total |
|---|---|---|---|---|---|
| 1983 | 11.3 | 1.2 | 6.4 | 6.4 | 25.6 |
| 1990 | 14.0 | 4.0 | 11.4 | 12.6 | 42.7 |
| 1995 | 14.4 | 7.8 | 14.3 | 17.1 | 53.6 |
| 2001 | 15.8 | 15.2 | 20.9 | 27.7 | 79.6 |
| 2005 | 18.5 | 18.5 | 29.0 | 37.0 | 103.5 |
| 2010[a)] | 19.5 | 19.5 | 37.0 | 50.0 | 126.0 |

[a)]Estimated.

Figure 39. World Polyolefin Production. (Kaminsky, 2008).

Personal research thus far has been an investigation of graphic control in shrink-wrap film. The purpose was to first understand distortion, its significance in various fields, especially packaging design, and to investigate the ability of the Human Visual System (HVS) to recognize it. Distortion may be here defined primarily as a lack of proportionality in an image resulting from defects, a loss of natural, normal, or original shape or condition. An understanding of polyolefins and their product films allows better control of graphics during printing and shrinking.

Polyolefins have become a pivotal material whose influence is far reaching and growing. Advances in production have allowed versatile applications and new pressures on natural resources will drive ingenuity and cost saving measures in the coming years.

REFERENCES


Aaker, J. L., Benet-Martinez, V., & Garolera, J. (2001). Consumption Symbols as Carriers of Culture: A Study of Japanese and Spanish Brand Personality Constructs. *Journal of Personality and Social Psychology,* 81 (3), 492–508. doi:10.1037//0022-35l4.g 1.3.492

AC Nielsen Consumer Study. (2010). *Shrink Films: An alternative to traditional labels*. Eastman Chemical Company. Retrieved 04.23.15 from <http://www.eastman.com/Brands/Eastman_Embrace/Media_Center/Pages/Literature_Center.aspx>

Agrawala, M., Zorin, D., & Munzner, T. (2000). Artistic multiprojection rendering. Paper presented at *the Proceedings of the Eurographics Workshop on Rendering Techniques,* 125-136.

Alfaro-Rodriguez, H., Angulo, O., & O'Mahony, M. (2007). Be your own placebo: A double paired preference test approach for establishing expected frequencies. *Food Quality and Preference,* 18, 353–361. doi:10.1016/j.foodqual.2006.02.009.

Alter, A. L., Oppenheimer, D. M., & Zemla, J. C. (2010). Missing the trees for the forest: A construal level account of the illusion of explanatory depth. *Journal of Personality and Social Psycholog*y, 99, 436–451. doi:10.1037/a0020218.

American Psychological Association (2010). Ethical principles of psychologists and code of conduct. Washington, DC: American Psychological Association.

Aronoff, J., Barclay, A. M., & Stevenson, L. A. (1988). The recognition of threatening facial stimuli. *Journal of Personality and Social Psychology*, 54(4), 647–655. doi.org/10.1037/0022-3514.54.4.647.

Aronoff, J., Woike, B. A., & Hayam, L. M. (1992). Which are the stimuli in facial displays of anger and happiness? Configurational bases of emotion recognition. *Journal of Personality and Social Psychology*, 62(6), 1050–1066. doi.org/10.1037/0022-3514.62.6.1050.

ASTM E1808-96(2009), Standard Guide for Designing and Conducting Visual Experiments, ASTM International, West Conshohocken, PA, 2009, www.astm.org.

ASTM E1432-04(2011), Standard Practice for Defining and Calculating Individual and Group Sensory Thresholds from Forced-Choice Data Sets of Intermediate Size, ASTM International, West Conshohocken, PA, 2011, www.astm.org.

Au, R. K., Ono, F., & Watanabe, K. (2013). Spatial distortion induced by imperceptible visual stimuli. *Consciousness And Cognition.* 22(1) 99-110. doi:10.1016/j.concog.2012.11.010.

Aydinoglu, N. Z., & Cian, L. (2014). Show me the product, show me the model: Effect of picture type on attitudes toward advertising. *Journal of Consumer Psychology,* 24(4) 506-519. doi.org/10.1016/j.jcps.2014.04.002.

Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science* 17(8), 645–648. doi:10.1111/j.1467-9280.2006.01759.x.

Bar, M., & Neta, M. (2007). Visual elements of subjective preference modulate amygdala activation. Neuropsychologia, 45, 2191–2200. doi:10.1016/j.neuropsychologia.2007.03.008.

Becker, Doug. (2014, February 19 & July 31). Personal correspondence with Fort Dearborn Company. [Telephone Interviews]

Bell, J., Forsyth, M., Badcock, D. R., & Kingdom, F. A. A. (2014). Global shape processing involves feature-selective and feature-agnostic coding mechanisms. *Journal of Vision*. 14(11), 12, 1-14. doi:10.1167/14.11.12.

Berinsky, A. J., Huber, G. A., & Lenz, G. A. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351–368. doi:10.1093/pan/mpr057.

Birnbaum, M. H. (2000). Psychological experiments on the Internet. San Diego, CA: Academic Press. Print.

Bloch, P. H., Brunel, F. F., & Arnold, T. J. (2003). Individual Differences in the Centrality of Visual Product Aesthetics: Concept and Measurement. *Journal of Consumer Research,* 29(4), 551–65.

Braun, V., Rogeaux, M., Schneid, N., O'Mahony, M., Rousseau, B. (2004). Corroborating the 2-AFC and 2-AC Thurstonian models using both a model system and sparkling water. *Food Quality and Preference*, 15, 501–507. doi:10.1016/j.foodqual.2003.10.002.

Bridgeman, B., Lennon, M., & Jackenthal, A. (2003) Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Test Performance. *Applied Measurement in Education*. 16(3), 191-205. doi:10.1207/S15324818AME1603_2.

Budweiser's 'Bowtie Shape' Can. (2013). *The Dieline.* Retrieved 04.23.15 from <http://www.thedieline.com/blog/2013/4/23/budweisers-bowtie-shape-can.html>

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6(1), 3–5. doi:10.1177/1745691610393980.

Busemeyer, J. R., Barkan, R., Mehta, S., & Chaturvedi. A. (2007). Context effects and models of preferential choice: implications for consumer behavior. *Marketing Theory,* 7(1), 39-58. doi: 10.1177/1470593107073844.

Cass, Jacob. (2009). *Vital Tips for Effective Logo Design.* Retrieved 02.08.15 from <http://www.smashingmagazine.com/2009/08/26/vital-tips-for-effective-logo-design/>

Cheskin, Louis. (1951). *Color For Profit.* Liveright Pubishing, NY, NY. Print.

Chou, M. C. & Wang, R. W. Y. (2011). Differentiation in the arched surface of packaging: Its influence on the findability of logo typography displays. *Displays,* 32(1), 24-34. doi:10.1016/j.displa.2010.09.006.

Chou, M. C. & Wang, R. W. Y. (2012). Displayability: An assessment of differentiation design for the findability of bottle packaging. *Displays*, 33(3), 146-156. doi:10.1016/j.displa.2012.06.003.

Christensen, R. H. B., Lee, H-S., & Brockhoff, P. B. (2012). Estimation of the Thurstonian model for the 2-AC protocol. *Food Quality* and Preference. 24(1), 119–128. doi:10.1016/j.foodqual.2011.10.005.

Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2012). From perception to conception: How meaningful objects are processed over time. *Cerebral Cortex,* 23(1), 187-197. doi:10.1093/cercor/bhs002.

Contour Bottle History. (2015). *Coca-Cola Great Britain.* Retrieved 04.23.15 from <http://www.coca-cola.co.uk/packages/history/contour-bottle-history/>

Detenber, B., and Reeves, B. (1996). A bio-informational theory of emotion: Motion and image size effects on viewers. *Journal of Communication,* 46(3), 66-84. doi:10.1111/j.1460-2466.1996.tb01489.x.

Djurovic, Vladimir. (2009). Brand translation: packaging design differences between China and the west. Retrieved 01.24.15 from <http://www.articledashboard.com/Article/Brand- Translation–Packaging-Design-Differences-between-China-and-the-West/ 874994>

Doyle, Mona. (1996). *Packaging Strategy: Winning the Consumer*. Technomic Publishing Company, Pennsylvania. Print.

Duncanson, J. P., & Williams, A. D. (1973). Video conferencing: Reactions of users. *Human Factors,* 15(5), 471-485. doi:10.1177/001872087301500504.

Economist Newspaper, The. (2012). America's Bounty: Gas Works. Retrieved 11.12.13 from <http://www.economist.com/node/21558459>

Elam, Kimberly. (2001). Geometry of design: Studies in proportion and composition. Princeton Architectural Press.

Eriksson, K., & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making*, 53(3), 159–163.

Fang, H., & Hart, J. C. (2007). Detail preserving shape deformation in image editing. Paper presented at the *ACM Transactions on Graphics (TOG),* 26(3), 12. doi:10.1145/1276377.1276392.

Fischer, Joanna. (2012). Determining the proper point size for display type on packaging. (M.S., Clemson University). *ProQuest Dissertations and Theses,* Retrieved from <http://search.proquest.com.libproxy.clemson.edu/docview/1039262740?accountid=6167> (1039262740).

Garber, Lawrence, L. (1995). The package appearance in choice. *Advances in Consumer Research,* 22(1) 653–660.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, 18(1), 23-30. doi:10.3758/BF03202642.

Gardner, Rick M. (1996). Methodological issues in assessment of the perceptual component of body image disturbance. *British Journal of Psychology*. 87(2), 327-337. doi:10.1111/j.2044-8295.1996.tb02593.x.

Geistfeld, L. V., Sproles, G. B., & Badenhop, S. B. (1977). The Concept and Measurement of a Hierarchy of Product Characteristics. *Advances in Consumer Research,* 4, 302–307.

Gescheider, George A. (1997). *Psychophysics: the fundamentals* (3rd ed.). Lawrence Erlbaum Associates. Print.

Giles, Geoff A. (2000). *Design and technology of packaging decoration for the consumer market.* Volume 1 of Sheffield packaging technology. CRC Press. Boca Raton, FL. Print.

Gobé, Marc. (2001) *Emotional Branding: The New Paradigm for Connecting Brands to People*. Allworth Press, New York. Print.

Godin, Seth. (2003) *Purple Cow: Transform Your Business by Being Remarkable*. Penguin Group. New York. Print.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O.P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, 59, 93–104.

Grabe, M., Lombard, M., Reich, R., Bracken, C., Ditton, T. (1999). The role of screen size in viewer experiences of media content. *Visual Communication Quarterly*, 6(2), 4-9. doi:10.1080/15551399909363403.

Grahl, A., Greiner, U., & Walla, P. (2012). Bottle Shape Elicits Gender-Specific Emotion: A Startle Reflex Modulation Study. *Psychology* 3(7), 548-554. doi:10.4236/psych.2012.37081.

Gray, J. R., Braver, T. S., & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences of the USA,* 99(6), 4115-4120. doi:10.1073/pnas.062381899.

Green, D.M. & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley. Print.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, 27, 649-677. doi:10.1146/annurev.neuro.27.070203.144220.

Hancock, P. J., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory & Cognition,* 24(1), 26-40. doi:10.3758/BF03197270.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psycho*logy, 54(6), 917-924. doi:10.1037/0022-3514.54.6.917.

Hekkert, P., & Leder, H. (2008). Chapter 10, Product aesthetics. *Product Experience*. Amsterdam: Elsevier. Print.

Henderson, P., Cote, J. A., Leong, S. M., & Schmitt, B. (2003) Building strong brands in Asia: Selecting the visual components of image to maximize brand strength. *International Journal of Research in Marketing*, 20(4), 297-313. doi:10.1016/j.ijresmar.2003.03.001.

Hertenstein, J. H., Platt, M. B., & Veryzer, R. W. (2013). What is "good design"?: An investigation of the complexity and structure of design. *Design Management Journal,* 8(1), 8-21. doi:10.1111/dmj.12000.

Holbrook, M. B., & Schindler, R. M. (1994). Age, sex, and attitude toward the past as predictors of consumers' aesthetic tastes for cultural products. *Journal of Marketing Research,* 31(3), 412-422. doi:10.2307/3152228.

Horton, J., and Chilton, L. (2010). The labor economics of paid crowdsourcing. Proceedings of the 11th ACM Conference on Electronic Commerce, Cambridge, MA. 209-218. doi:10.1145/1807342.1807376.

Iachini, S., & Giusberti, F. (1996). Metric Aspects of Mental Images. *Perceptual and Motor Skills*, 83(3), 1243-1250. doi:10.2466/pms.1996.83.3f.1243.

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? Paper presented at the *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on,* 145-152.

Jesty, L. C. (1958). The relation between picture size, viewing distance, and picture quality. *Proceedings of the IEE,* 105B(23), 425-439. doi:10.1049/pi-b-1.1958.0320.

Kanwisher, N., Chun, M. M., McDermott, J., & Ledden, P. J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research,* 5(2), 55-67. doi:10.1016/S0926-6410(96)00041-9.

Kam, C. D., Wilking, J. R., & Zechmeister, E. J. (2007). Beyond the 'narrow data base': Another convenience sample for experimental research. *Political Behavior,* 29(4), 415–440. doi:10.1007/s11109-007-9037-6.

Kaminsky, Walter. (2008). Trends in Polyolefin Chemistry. *Macromolecular Chemistry and Physics*, 209(5), 459-466. doi:10.1002/macp.200700575.

Karni, Z., Freedman, D., & Gotsman, C. (2009). Energy-based image deformation. *Computer Graphics Forum, 28*(5), 1257-1268. doi:10.1111/j.1467-8659.2009.01503.x.

Karunasekera, S. A., & Kingsbury, N. G. (1995). A distortion measure for blocking artifacts in images based on human visual sensitivity. *Image Processing, IEEE Transactions on,* 4(6), 713-724. doi:10.1109/83.388074.

Keller, Kevin L. (2003). Brand Synthesis: The Multidimensionality of Brand Knowledge. *Journal of Consumer Research*, 29(4), 595–600. doi:10.1086/346254.

Kim, T. (1996). The memory and persuasion effects of presence in television advertisement processing. Doctoral dissertation. University of North Carolina, Chapel Hill.

King, Laura. (2014). *The Science of Psychology: An Appreciative View*. 3rd Edition. McGraw-Hill. Print.

Kotler, Philip. (1997). *Marketing Management: Analysis, Planning, Implementation, and Control*. 9th Edition. Prentice Hall. Upper Saddle River, NJ. Print.

Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. H. Birnbaum (Ed.), Psychological experiments on the Internet (pp. 35–60). New York: Academic Press. doi:10.1016/B978-012099980-4/50003-4.

Kunst-Wilson, W. R. & Zajonc, R.B. (1980). Affective Discrimination of Stimuli that Cannot Be Recognized. *Science,* 207(4430), 557–58.

Larson, E., & Chandler, D. (2010). Most apparent distortion: Full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging,* 19(1), 011006. doi:10.1117/1.3267105.

Lewicki, Pawel (1986). *Nonconscious Social Information Processing*. New York: Academic Press, Inc. Print.

Li, Y., Nam, C., Shadden, B., Johsnon, S. (2010). A P300-Based Brain–Computer Interface: Effects of Interface Type and Screen Size. *International Journal of Human-Computer Interaction*, 27(1), 52-68. doi:10.1080/10447318.2011.535753.

Lidwell, W., Holden, K., & Butler, J. (2010). *Universal principles of design: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design.* Rockport Publishing.

Lin, W., Dong, L., & Xue, P. (2005). Visual distortion gauge based on discrimination of noticeable contrast changes. *Circuits and Systems for Video Technology, IEEE Transactions on,* 15(7), 900-909. doi:10.1109/TCSVT.2005.848345.

Lin, W., & Kuo, C. (2011). Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation,* 22(4), 297-312. doi:10.1016/j.jvcir.2011.01.005.

Loken, B. & Ward, J. (1990). Alternative Approaches to Understanding the Determinants of Typicality. *Journal of Consumer Research,* 17(2), 111–26. doi:10.1086/208542.

Lombard, Matthew. (1995). Direct responses to people on the screen: Television and personal space. *Communication Research,* 22(3), 288-324. doi:10.1177/009365095022003002.

Lombard, M., Ditton, T. B., Grabe, M. E., & Reich, R. D. (1997). The role of screen size in viewer responses to television. *Communication Reports,* 10(1), 95-106. doi:10.1080/08934219709367663.

Lund, Arnold. M. (1993). The influence of video image size and resolution on viewing-distance preferences. *SMPTE Journal,* 102(5), 406-415. doi:10.5594/J15915.

Marchisano C., Lim J. Y., Cho H. S., Suh D. S., Jeon S. Y., Kim K. O., & O'Mahony, M. (2003). Consumers report preferences when they should not: A cross-cultural study. *Journal of Sensory Studies*. 18(6), 487–516. doi:10.1111/j.1745-459X.2003.tb00402.x.

McCulloch, Charles & Searle, Shayle. (2001). *Generalized, Linear, and Mixed Models*. 1st Edition. Chichester: John Wiley & Sons. Print.

Meilgaard, M. C., Civille, G. V., & Carr, B. T. (1999). *Sensory Evaluation Techniques.* 3rd Edition. CRC Press. Print.

Morris, Max D. (2011). *Design of Experiments: An Introduction Based on Linear Models*. Chapman & Hall/CRC. Boca Raton, FL.

Nathan, J. G., Anderson, D. R., Field, D. E., & Collins, P. (1985). Television viewing at home: Distances and visual angles of children and adults. *Human Factors,* 27(4), 467-476. doi: 10.1177/001872088502700410.

Ochsner, K. N., & Phelps, E. (2007). Emerging perspectives on emotion-cognition interactions. *Trends in Cognitive Sciences*. 11(8), 317-318. doi:10.1016/j.tics.2007.06.008.

Ohtani, T., & Mitshuhashi, T. (1973). Picture quality of a high definition television system. *TEBS, 3-4.*

Orban, G. A., Dupont, P., Vogels, R., De Bruyn, B., Bormans, G., & Mortelmans, L. (1996). Task dependency of visual processing in the human visual system. *Behavioural Brain Research,* 76(2), 215-223. doi:10.1016/0166-4328(95)00195-6

Orth, U. R., & Malkewitz, K. (2008). Holistic Package Design and Consumer Brand Impressions. *Journal of Marketing* 72(3), 64-81. doi.org/10.1509/jmkg.72.3.64.

Orth, U. R., Campana, D., & Malkewitz, K. (2010). Formation of consumer price expectation based on package design, attractive and quality routes. *Journal of Marketing Theory and Practice*. 18(1), 23-40. doi:10.2753/MTP1069-6679180102.

Oyama, T., Agostini, T., Kamada, A., Marcovic, S., Osaka, E., Sakura, S., Sarmany-Schuller, S. & Sarris, V. (2008). Similarities in form symbolism among various languages and geographical regions. *Psychologia*, 51(3), 170–184. doi:10.2117/psysoc.2008.170.

O'Mahony, M., & Rousseau, B. (2002). Discrimination testing: A Few Ideas, Old and New. *Food Quality and Preference*, 14(2), 157-164. doi:10.1016/S0950-3293(02)00109-X.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419.

Pappas, T. N., Safranek, R. J., & Chen, J. (2000). Perceptual criteria for image quality evaluation. *Handbook of Image and Video Processing.* 669-684. doi:10.1016/B978-012119792-6/50118-2.

Pentland, Alex. (1988). Shape information from shading: A theory about human perception. *Spatial Vision,* 4(2), 165-182. doi:10.1163/156856889X00103.

Plastemart.com. (2011). *Global Polyolefin Market Changes*. Retrieved 11.12.13 from <http://www.plastemart.com/Plastic-Technical-Article.asp?LiteratureID=1714&Paper=global-polyolefin-market-sees-dramatic-changes>

Pontin, Jason. (2007). *Artificial intelligence: With help from the humans.* The New York Times. Retrieved 11.05.14 from <http://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html>

Prescott, J., Norris, L., Kunst, M., & Kim, S. (2005). Estimating a "Consumer Rejection Threshold" for Cork Taint in White Wine. Food Quality and Preference. 16(4), 345-349. doi:10.1016/j.foodqual.2004.05.010.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *The Journal of Neuroscience*. 16(16), 5205-5215.

Raghubir, P., & Greenleaf, E. A. (2006). Ratios in proportion: what should the shape of the package be?, *Journal of Marketing,* 70(2), 95–107. doi:10.1509/jmkg.70.2.95.

Rachman, S. & Cuk, M. (1992). Fearful distortions. *Behavior Research and Therapy*. 30(6), 583-589. doi:10.1016/0005-7967(92)90003-Y.

Rettie, R. & Brewer, C. (2000). The verbal and visual components of package design, *Journal of Product and Brand Management*, 9(1), 56–70. doi:10.1108/10610420010316339.

Reviewing Research on Mechanical Turk. (2010). *IRB Advisor*. 10 (12), 140. Retrieved 11.05.14 from ProQuest database.

Robertson, Gordon L. (2006). *Food Packaging, Principles and Practice.* 2nd Edition. Taylor & Francis. Boca Raton, FL. Print.

Roncarelli, S. & Ellicott, C. (2010). *Packaging Essentials: 100 design principles for creating packages*. Beverly, Mass. Rockport Publishers. Print.

Rouse, Steven. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior,* 43, 304–307. doi:10.1016/j.chb.2014.11.004.

Sage Publications & Goldstein, Sage eReference (Online service). (2010). *Encyclopedia of Perception*. Thousand Oaks, Calif.: Sage Publications, Inc.

Seal, Barry. (2009). *The importance of good packaging design has never been so great.* Retrieved 01.24.15 from <http://www.creativematch.co.uk/viewNews/?97019>

Schaefer, S., McPhail, T., & Warren, J. (2006). Image deformation using moving least squares. Paper presented at the *ACM Transactions on Graphics (TOG),* 25(3) 533-540.

Schneider, Abigail. (2014). Message in a bottle: What anthropomorphized packaging design tells consumers about themselves (Doctorial dissertation). Retrieved from ProQuest. (3621406)

Sheikh, H. R., & Bovik, A. C. (2006). Image information and visual quality. *Image Processing, IEEE Transactions on*, 15(2), 430-444. doi:10.1109/TIP.2005.859378.

Soares, J. & McKenna, T. F. L. (2012). *Polyolefin Reaction Engineering.* 1st Edition. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. Print.

Smith, P. R. & Taylor, J. (2004). *Marketing Communications: An Integrated Approach*. 4th Edition. Kogan Page, London. Print.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520-522.

Tonzani, Stefano. (2012). The Renaissance of Polyolefins. *Journal of Applied Polymer Science*, 127(2), 837. doi:10.1002/app.38393.

Underwood, Robert L. (2003). The Communicative Power of Product Packaging: Creating Brand Identity via Lived and Mediated Experience. *Journal of Marketing Theory and Practice,* 11(1), 62–76.

U.S. Patent Office. (2013). *Heat-Shrinkable Film*. Retrieved 11.12.13 from <http://pdfpiw.uspto.gov/.piw?docid=02876067&SectionNum=4&IDKey=E18FACC5AD96&HomeUrl=http://patft.uspto.gov/netacgi/nph-Parser?Sect2=PTO1%2526Sect2=HITOFF%2526p=1%2526u=%25252Fnetahtml%25252FPTO%25252Fsearch-bool.html%2526r=1%2526f=G%2526l=50%2526d=PALL%2526S1=2876067.PN.%2526OS=PN/2876067%2526RS=PN/2876067>

Veryzer, Robert W. (1999). A Nonconscious Processing Explanation of Consumer Response to Product Design. *Psychology and Marketing,* 16(6), 497-522. doi:10.1002/(SICI)1520-6793(199909)16:63.0.CO;2-Z.

Wang, Z., & Bovik, A. C. (2002). A universal image quality index. *Signal Processing Letters, IEEE*, 9(3), 81-84. doi:10.1109/97.995823.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on,* 13(4), 600-612. doi:10.1109/TIP.2003.819861.

Wansink, B., Dyson, J. & von Ittersum, Koert. (2005). Shape of glass and amount of alcohol poured: comparative study of effect of practice and concentration. *Business Management Journal,* 331(7531), 1512–1514. doi:10.1136/bmj.331.7531.1512.

Walla, P., Hufnagl, B., Lehrner, J., Mayer, D., Lindinger, G., Deecke, L., & Lang, W. (2002). Evidence of conscious and subconscious olfactory information processing during word encoding: A magnetoencephalographic (MEG) study. *Cognitive Brain Research*, 14(3), 309-316. doi:10.1016/S0926-6410(02)00121-0.

Walla, P., Hufnagl, B., Lehrner, J., Mayer, D., Lindinger, G., Imhof, H., Deecke, L., & Lang, W. (2003). Olfaction and depth of word processing: A magnetoencephalographic (MEG) study. *Neuroimage,* 18(1), 104-116. doi:10.1006/nimg.2002.1324.

Walla, Peter. (2008). Olfaction and its dynamic influence on word and face processing: Cross-modal integration. *Progress in Neurobiology,* 84(2), 192-209. doi:10.1016/j.pneurobio.2007.10.005.

Wesp, R., Peckyno, A., McCall, S., & Peters, S. (2000). Object recognition may distort size perception. *Perceptual and Motor Skills*, 90(3), 803-809. doi:10.2466/PMS.90.3.803-809.

Westerink, J. H. D. M., & Roufs, J. A. J. (1989). Subjective image quality as a function of viewing distanc0e, resolution and picture size. *SMPTE Journal,* 98(2), 113-119. doi:10.5594/J02825.

Westerman, S. J., Gardner, P. H., Sutherland, E. J., White, T., Jordan, K., Watts, D., & Wells, S. (2012). Product Design: Preference for Rounded versus Angular Design Elements. *Psychology & Marketing.* 29(8), 595-605. doi:10.1002/mar.20546.

Westerman, S. J., Sutherland, E. J., Gardner, P. H., Baig, N., Critchley, C., Hickey, C., Mehigan, S., Soloway, A., & Zervos, Z. (2013). The design of consumer packaging: Effects of manipulations of shape, orientation, and alignment of graphical forms on consumers' assessments. *Food Quality and Preference.* 27(1), 8–17. doi:10.1016/j.foodqual.2012.05.007.

Winkielman, P., & Berridge, K. C. (2004). Unconscious emotion. *Current Directions in Psychological Science,* 13(3), 120-123. doi:10.1111/j.0963-7214.2004.00288.x.

Yamashita, W., Kanazawa, S., Yamaguchi, M. K. (2014). Tolerance of geometric distortions in infant's face recognition. *Infant Behavior & Deve*lopment, 37(1), 16-20. doi:10.1016/j.infbeh.2013.10.003.

Yap, Ewam. (2013). *BIG BRAND THEORY: Experimental Packaging Design on Behance.* Web Blog. Retrieved 11.05.14 from <https://www.behance.net/gallery/6546453/BIG-BRAND-THEORY-Experimental-Packaging-Design>

Young, Scott. (2006). Measuring success: Using consumer research to document the value of package design. *Design Management Review,* 17(2), 60–65. doi:10.1111/j.1948-7169.2006.tb00040.x.

Yu, D., Legge, G. E., Wagoner, G., & Chung, S. T. L. (2014). Sensory factors limiting horizontal and vertical visual span for letter recognition. *Journal of Vision,* 14(9), 3, 1-17. doi:10.1167/14.9.23.

Zorin, D., & Barr, A. H. (1995). Correction of geometric perceptual distortions in pictures. Paper presented at the *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques,* 257-264.

Xue, L., & Yen, C. C. (2007). Towards female preferences in design - A pilot study. *International Journal of Design,* 1(3), 11-27.