The use of immersive technologies to improve consumer testing: the impact of multiple immersion levels on data quality and panelist engagement for the evaluation of cookies under a preparation-based scenario

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By
Drew Hathaway
Graduate Program in Food Science and Technology

The Ohio State University
2015

Master's Examination Committee:
Dr. Christopher T. Simons, Advisor
Dr. Luis Rodriguez-Saona
Dr. Ken Lee
Abstract

Companies in the Food and Beverage industry invest large amounts of money into consumer sensory testing to support new product development. However, even when consumer data suggests that a product should succeed, only a small percentage of new food products survive once in the marketplace. The inability of traditional consumer testing methodologies to reliably predict consumers’ future purchase decisions is thought to be a major contributor of this high failure rate. Commonly used practices place consumers in isolated booths designed to control against the influence of non-product (e.g. environmental) factors. Such testing conditions lack meaningful contextual information important in forming consumers’ subsequent perceptions, hedonic assessments, and behavior towards a product. Consequently, these sterile environments and monotonous testing procedures often fail to adequately engage consumer panelists which can result in uninformative or misleading consumer data. Therefore, the lack of environmental context and panelist engagement under traditional consumer sensory testing practices are thought to be predominantly responsible for the inaccuracy of data collected using this approach.

The present study investigated the effect of utilizing immersive technologies to provide contextual information relevant to the preparation of cookies within a sensory testing environment on the resulting hedonic data and subjective assessment of engagement collected using a provided questionnaire. This was accomplished by
comparing results in which subjects evaluated the same cookie samples in a traditional sensory testing booth, an identical sensory booth complete with visual, auditory, and olfactory cues associated with cookie preparation, and a full immersion testing room complete with same three context cues on two separate occasions approximately three weeks apart. We hypothesized that improving the ecological validity of consumer hedonic testing by incorporating relevant visual, audio and olfactory cues through the use of immersive technologies would result in a more engaging testing experience and hedonic assessments that are more discriminating and reliable compared to those derived from traditional testing paradigms. We further hypothesized that the different levels of immersion utilized in the full immersion and mixed immersion environments during testing would differentially impact how much improvement was observed in the sensitivity, power and reliability of hedonic data and the subjective assessment of engagement when compared to results from the traditional sensory testing environment. Indeed, we found the presence of contextual information and the degree of immersion employed to have an effect on the panelists’ evaluations of the cookie samples and their subjective assessment of engagement. Results collected in the two immersive environments were found to be slightly more discriminating than the data collected in the traditional testing environment, particularly for the second session. The two immersive environments also proved to be more engaging to consumer panelists than the traditional environment which may have contributed to the higher average hedonic ratings given in these environments. Furthermore, the two immersive environments remained equally engaging during both sessions whereas the traditional testing environment was found to
be less engaging during the second session. This may have been partially responsible for the decrease in power observed from the first session to the second in the traditional environment compared to the immersive environments. Moreover, this finding suggests that the impact of immersive environments as designed presently, was enduring, at least over the approximately 3-week test period evaluated here. In general, the full immersion environment produced the best results out of the three environments. This along with similar findings from Bangcuyo et al., (2015) suggests that the utilization of an immersive testing room might represent the most optimal design of this testing practice to date. However, the mixed immersion environment still showed that it might have the potential to be a viable alternative to a full immersion lab as it generally produced hedonic data of a higher quality than the traditional environment albeit not quite as compelling as the full immersion environment. Adoption of the mixed immersion methodology by the food industry would only require companies to slightly modify their current sensory testing booths to contain visual, audio, and olfactory components relevant to the product being evaluated while maintaining their normal high throughput ability, thus making it the more ideal design. The improvement in data quality observed in the two immersive environments containing preparation-based contextual information about cookies also suggests that implementation of this consumer testing methodology, or a similarly modeled one, might be beneficial to the large number of companies who produce food products meant to be prepared and consumed at home. Furthermore, these results suggest that methodological changes to current consumer testing practices have
the potential to improve the quality of hedonic data providing food manufacturers significant savings on product development costs and failed product launches.
Acknowledgements

I want to thank my family and friends for always supporting me and encouraging me to pursue what interests me throughout my life. Additionally I want to give a special thanks to Dr. Simons for giving me the opportunity to further my education at The Ohio State University. This work was supported by the USDA National Institute of Food and Agriculture, Hatch project [233590].
Vita

2014 ............................................B.S. Food Science and Technology,
   The Ohio State University

2014 to present ............................................Graduate Student, Department of Food
   Science and Technology, The Ohio State
   University

Fields of Study

Major Field: Food Science and Technology
**Table of Contents**

Abstract ............................................................................................................................................... ii

Acknowledgements ............................................................................................................................... vi

Vita........................................................................................................................................................ vii

List of Tables ......................................................................................................................................... xi

List of Figures ......................................................................................................................................... xii

Chapter 1: Literature Review ............................................................................................................... 1

1.1 Consumer Sensory Testing ............................................................................................................. 1

1.2 Context ........................................................................................................................................... 3

1.3 Engagement ................................................................................................................................... 5

1.4 Immersive Technologies ................................................................................................................. 6

Chapter 2: Introduction ......................................................................................................................... 11

Chapter 3: Material & Methods ........................................................................................................... 14

3.1 Participants .................................................................................................................................... 14

3.2 Stimuli ........................................................................................................................................... 14

3.3 Procedure ..................................................................................................................................... 16
List of Tables

Table 1. Average hedonic scores for cookies separated by testing environment and replicate................................................................. 67

Table 2. Average engagement scores for each dimension separated by testing environment and replicate......................................................... 68
List of Figures

Figure 1. Cookie liking results obtained from panelists evaluating samples in each of the three conditions. A. Combined average hedonic ratings across cookies separated by condition and replicate. B. Average hedonic ratings for each cookie separated by condition and replicate .......................................................... 27

Figure 2. Average hedonic score for each cookie type across replicates separated by testing environment ............................................................................................................. 31

Figure 3. Change in mean liking scores for each cookie in each testing environment..... 32

Figure 4. Average change in combined mean liking scores combined for all cookies in each testing environment ............................................................................................................. 33

Figure 5. Subjective assessment of panelist engagement in each of the three testing environments collected using the provided questionnaire. A. Average ratings for engagement dimensions for each testing environment separated by replicate. B. Average total engagement score for each testing environment separated by replicate .......... 37

Figure 6. Testing environments .................................................................................. 65

Figure 7. Diagram outlining the physical setup of the three testing environments ....... 66
Chapter 1: Literature Review

1.1 Consumer Sensory Testing

Most companies in the Food and Beverage industry are continuously working to create and launch new products into the marketplace in order to gain competitive advantage and long term financial success while attempting to keep up with the constantly changing landscape of consumer trends and desires (Costa & Jongen 2006). But the number of new food products that succeed in the marketplace is far outweighed by those that ultimately end up failing with research finding that nearly 60% of all new product launches in the grocery sector fail within one year (Costa & Jongen 2010) and approximately 80% of all new food products fail eventually (Redmond 1995). As a consequence, billions of dollars are spent annually by the Food and Beverage industry on failed product launches (Stewart-Knox & Mitchell 2003). This not only means that manufacturers fail to see an adequate return on investment but it also contributes to higher prices and less affordable food for consumers (Bangcuyo, Simth, Zumach, Pierce, Guttman, & Simons 2015). The high failure rate of new food product launches is likely a result of multiple factors which could include but is not limited to technical limitations, poor marketing and/or pricing strategies (Bangcuyo et al., 2015). One of the factors believed to be a major contributor to this high failure rate is the inability of current consumer sensory testing methodologies to reliably predict consumers’ future purchase
decisions (Deliza & MacFie 1996; Koster & Mojet 2007; Rosas-Nexticapa, Angulo & O’Mahony 2005). Most food companies regularly perform consumer acceptance tests on new products before they are launched into the marketplace but the results are only meaningful if they can accurately predict the behavior of the consumers in the real world (Weiss, O’Mahony, & Wichchukit 2010). However there is often a lack of agreement between what consumers report liking or would be likely to choose or buy during consumer sensory tests with what they actually end up choosing to purchase once leaving the testing situation (Weiss et al., 2010). Most commonly used consumer testing methodologies force panelists to evaluate product samples while they are in isolated sensory booths intentionally designed to lack any contextual information. This sterile environment is utilized in order to prevent any confounding non-product factors from influencing a consumers’ assessment (Bangcuyo et al., 2015). However, such testing practices have been shown to result in hedonic data that often have little predictive power or may even be misleading (Bangcuyo et al., 2015). This lack of ecological validity presents a serious issue since companies use these data to make important business decisions such as selecting which prototype to launch into the marketplace as a new product. Some food manufacturers recognize the limitations of current consumer sensory testing practices and in an attempt to overcome them have explored alternative ways of conducting consumer tests. The two most common forms of these alternatives are home-use and on-site tests because they enable participants to evaluate the product samples in an environment of typical use in an attempt to improve the ecological validity of the resulting data. (Bangcuyo et al., 2015). While this methodology is ideal in theory, there
have been multiple limitations found to be associated with the execution of these types of tests (Meilgaard, Civille, & Carr 1999). In particular, they tend to be more expensive, time consuming, difficult to execute, and lack experimenter oversight potentially allowing for a variety of confounding non-product factors to influence the results (Meilgaard et al., 1999). In addition, these tests are only recommended for evaluating a very small number of samples—three or less—to maintain the natural use situation that they are designed for (Meilgaard et al., 1999). This makes these testing methodologies inapplicable for commonly used multi-sample tests such as product optimization or category review (Meilgaard et al., 1999). As a consequence of these limitations, these types of tests are only ever conducted on a select few prototypes which were likely selected previously using conventional sensory testing methods (Bangcuyo et al., 2015). Therefore, there is a serious need for the development of new consumer sensory testing methodologies capable of producing more ecologically valid hedonic data while maintaining logistical feasibility.

1.2 Context

In real-world settings, people consume food products in the presence of a composite of contextual information. These conditions contain sources of dynamic visual, auditory, and olfactory stimuli that are processed simultaneously by the brain to create complex contextual cues about the environment. As this contextual information is continually processed it can impact a consumer’s subsequent expectations, perceptions, and behavior about a product or situation (Bangcuyo et al., 2015). Previous research has
demonstrated that the presence of contextual information, whether simplistic or complex, can completely change a consumers’ resulting hedonic assessment of a product (Delarue & Boutrolle 2010; Bangcuyo et al., 2015). Contextual information about the product itself such as the product name (Herz & Von Clef 2001; Wansink, Van Ittersum & Painter 2005), packaging (Ares & Deliza 2012), nutritional information (Liem, Aydin, & Zandstra 2012), color (Morrot, Brochet & Dubourdieu 2001; Pangborn, Berg & Hansen 1963), temperature (Zellner, Stewart, Rozin & Brown 1988), and aroma (Yeomans 2006) of food have all been shown to have a profound influence on a consumers’ perception and liking of the product. Many environmental factors have also been shown to influence consumer liking and food-related behaviors such as ambiance (Delarue & Boutrolle 2010; Stroebele & De Castro 2004), eating location (Bell, Meiselman, Pierson & Reeve 1994; Delarue & Boutrolle 2010), number of people present (Sommer & Steele 1997), drinking vessel (Raudenbush, Meyer, Eppich, Corley, & Petterson 2002), plate weight (Piqueras-Fiszman, Harrar, Alcaide & Spence 2011), ambient temperatures (Westerterp-Platenga 1999), sounds (Ferber & Cabanac 1987) and lighting (Kasof 2002).

Despite all of these data illustrating the importance of context cues during real-life product evaluation, current consumer sensory testing methodologies typically have subjects evaluate products in small isolated booths intentionally devoid of contextual information. Under such testing conditions, important contextual information is absent resulting in panelists evaluating a product in an environment that is nothing like where or how they would normally consume the product in real-life circumstances (Bangcuyo et al., 2015). As a consequence, much of the sensory data gathered by food companies lacks
ecological validity forcing them to make important decisions based on data that may or may not be truly representative of the panelists’ preference or liking.

1.3 Engagement

In addition to the absence of context, lack of panelist engagement during traditional consumer sensory tests is also believed to be a major contributor to the poor quality and predictive ability of hedonic data (Bangcuyo et al., 2015). Creating engaging user experiences is essential to the design of any system requiring interaction (O’Brien & Toms 2010). Understanding the construct of engagement and how to improve and/or manipulate it is a topic with a significant amount of research due to its applicability across a wide variety of fields (O’Brien & Toms 2010). Other research has focused on identifying and understanding the underlying concepts that make up the complex construct of engagement and assessing the most optimal methods for measuring it (O’Brien & Toms 2010, O’Brien 2008). It proposed that focused attention and user motivation are two of the most important elements driving consumer engagement (O’Brien 2008). This notion was based on the general understanding that participants have to be focused in order to be engaged and that people are more motivated to participate in enjoyable, engaging experiences (O’Brien 2008). Other important factors contributing to the level of engagement during testing include aesthetics and novelty (O’Brien 2008). Aesthetically pleasing environments are understood to increase general arousal. Similarly, humans tend to be curious and seek variety and novelty in their daily
encounters. Therefore the inclusion of factors found to be aesthetically pleasing and novel should improve a person’s level of felt engagement (O’Brien & Toms 2010).

This presents an issue in the field of traditional consumer sensory testing as current methodologies often fail to adequately engage consumer panelists as a consequence of monotonous procedures and sterile testing environments (Bangcuyo et al., 2015). As a result many participants fail to focus intently on the evaluation task at hand and, instead, provide the same (or nearly the same) hedonic score for all samples (Koster 2009). These non-discriminating panelists significantly limit the utility of the resulting hedonic data (Bangcuyo et al., 2015).

1.4 Immersive Technologies

The emergence of new and constantly improving immersive technologies in recent history has brought with it the possibility of creating simulations meant to imitate real-life scenarios (Witmer & Singer 1998). This has allowed for a fundamentally new approach to developing research methodologies meant to explore consumer behavior. Immersive technologies exist in many forms, and the design in which they are utilized can affect a participants’ resulting sense of immersion (Dede 2009). One example is the utilization of head-mounted displays to digitally replicate real-world environments (Dede 2009). When this approach is utilized properly it can be an effective way of inducing user immersion but presently there are still many limitations associated with the technology. This technology is especially limited under testing scenarios meant to explore a users’ interaction with tangible objects outside of the portrayed simulation, such as the
consumption of food samples. Another approach is the interaction of a user with a computer generated three-dimensional online environment (Daugherty, Li & Biocca 2005). The business sector has already utilized this approach extensively to identify how a variety of extrinsic product properties such as packaging, shelf placement, pricing, and advertising can influence a consumers’ resulting purchasing behavior (for review see Daugherty et al., 2005). However, the present study explored the utilization of immersive technologies to create immersive virtual reality testing rooms specific to a particular scenario through the manipulation of contextual information. Previous research has explored how differences in the design of these immersive virtual reality testing rooms can affect consumers’ resulting behaviors. One such study by Sester et al., (2013) explored the effect of context on drink selection through the use of two different virtual bar-like environments created through the manipulation of visual, audio and environmental factors (e.g. furniture design). They found that the consumers’ drink choice changed based on the difference in ambiances portrayed in the two different virtual bars illustrating that context can influence consumer behavior (Sester et al., 2013).

Other research has been conducted to identify and measure the underlying components that make up an effective immersive environment and to evaluate how these technologies can best be utilized as research tools (Witmer & Singer 1998). Such efforts identified a participant’s sense of presence to be a key component of any immersive environment found to be effective in creating engaging experiences (Witmer & Singer 1998). This work also resulted in questions designed to assess the incidence of attributes thought to comprise the construct of user presence within a virtual environment including
perceived usability, sensory awareness, felt involvement, realism, distraction, and immersion (Witmer & Singer 1998). They proposed that the sense of presence in an immersive environment requires directed attention and is a result of the complex interaction between sensory stimulation, environmental factors that encourage involvement and enable immersion, and internal tendencies to become involved when engaged (Witmer & Singer 1998). In general, as users’ increase their focus on stimuli within a virtual environment they become more involved in the experience and immersed in the environment (Witmer & Singer 1998). When interacting with a virtual environment, design factors causing poor usability can impede a user’s ability to complete a task, decreasing engagement and therefore involvement (Witmer & Singer 1998). They further postulated that a participant’s sense of presence and therefore engagement can be improved within a virtual testing environment via the inclusion of factors that replicate real-world experiences and activate multiple sensory pathways while simultaneously minimizing factors found to be distracting (Witmer & Singer 1998).

Since lack of environmental context and panelist engagement under traditional consumer sensory testing practices are thought to be predominantly responsible for the inaccuracy of data collected using this approach, the utilization of a virtual testing environment was proposed by Bangcuyo et al., (2015) as a potential way of overcoming these limitations. This study was one of the first to investigate the use of immersive technologies in consumer hedonic testing of food products. As in the present study, they sought to better understand how the inclusion of relevant contextual clues within a sensory testing environment might impact the consumers’ engagement and resulting
hedonic data quality when compared to a traditional testing environment devoid of context (Bangcuyo et al., 2015). They did so by having the same panelists evaluate five coffee samples in a traditional sensory testing environment as well as in an immersive environment meant to depict a virtual coffeehouse by containing visual, auditory, and olfactory cues found regularly in a coffeehouse setting (Bangcuyo et al., 2015). The panelists also provided a subjective assessment of their engagement in each of the two environments via a provided questionnaire (Bangcuyo et al., 2015). Each panelist completed the same test on two separate occasions approximately 1-month apart so that the reliability of the hedonic data could be evaluated (Bangcuyo et al., 2015). This study found hedonic data collected in the virtual coffee house to be more discriminating and a more reliable predictor of future product preference versus data that were collected in the traditional testing environment devoid of contextual information (Bangcuyo et al., 2015). This was reflected by the fact that the preference order for the coffee samples only remained stable between sessions when it was evaluated in the virtual coffeehouse and by its ability to better resolve product differences by generating a much larger number of significant differences in the average liking scores between coffee samples for both sessions. They additionally found consumers to be more engaged in the testing when evaluating coffees in the virtual coffeehouse, which was believed to have been partially responsible for the improved quality of hedonic data collected in this environment versus the traditional one (Bangcuyo et al., 2015). These results suggest that during product evaluations, consumers unconsciously process contextual information about their environment simultaneously with intrinsic product attributes to arrive at their resulting
hedonic assessments (Bangcuyo et al., 2015). This study implies that incorporating immersive technologies in the field of consumer sensory testing can improve the ecological validity of testing conditions by restoring relevant contextual information resulting in data that are more discriminating and reliable (Bangcuyo et al., 2015). The immersive methodology also proved to be significantly more powerful than traditional consumer sensory testing, producing discriminating and reliable data with only 50 participants (Bangcuyo et al., 2015). Since traditional consumer sensory tests normally field hundreds of panelists this methodology presents the potential to significantly lower the execution costs associated with such testing (Bangcuyo et al., 2015). Therefore, these results suggest that the use of immersive environments in consumer sensory testing not only has the potential to improve the quality of hedonic data, but also might be able to provide companies with a significant savings opportunity on product development costs and failed launches (Bangcuyo et al., 2015). However, other than the study by Bangcuyo et al., (2015), the use of immersive environments in consumer acceptance testing is novel and further research is needed to validate these claims and further understand the extent of its potential application within the food industry.
Chapter 2: Introduction

A sensory study by Bangcuyo et al., (2015) found that the inclusion of relevant context cues within a virtual coffeehouse testing environment produced consumer hedonic data from five coffee samples that were more discriminating and reliable compared to data obtained by the same panelists in a traditional sensory testing environment devoid of context over two testing sessions. Additionally, consumers in the study reported feeling more engaged when evaluating coffees in the virtual coffeehouse versus the traditional environment. The success of Bangcuyo et al., (2015) at improving the quality of hedonic data and user engagement provides insight into the promising potential of utilizing immersive environments in consumer sensory testing as it represents one of many ways that immersive technologies can be implemented into the field.

Therefore, in the present research we sought to build upon the study by Bangcuyo et al., (2015) in two important ways. Firstly, since the previous study was based upon the scenario of a product purchased and consumed at a retail location we wanted to investigate if similar results would be produced for a home preparation/consumption scenario. These results would provide insight as to how the use of immersive technologies in consumer sensory testing could also be relevant for companies that produce food products designed to be prepared and consumed at home. This would represent a significant finding since these are the food companies who have traditionally
utilized consumer sensory testing. Secondly, in Bangcuyo et al., (2015), investigators only compared the effect of two environments, a virtual coffeehouse versus a traditional sensory testing environment, on the sensitivity, discriminability (testing power), and reliability of the resulting hedonic data. This presents a significant limitation for adoption of this methodology by the food industry because following the previous study’s example would mean that companies would have to create their own virtual reality testing environment. Using a virtual reality testing room for consumer sensory testing following the example of Bangcuyo et al., (2015) would also mean that companies would only be able to test one panelist at a time making the logistics and cost of this approach much less desirable to companies who are used to being able to test small groups of consumers at once.

Therefore, in this study we explored adding the same immersive elements used in a virtual testing environment to a traditional sensory testing booth to determine if the results would still show improved data quality and consumer engagement. If proven effective, then this would allow companies to easily modify their existing sensory testing booths to provide context cues congruent with the products being tested in order to improve consumer data while maintaining the same high through-put. Therefore, in the present study we had the same consumers evaluate cookie samples in three environments, a traditional sensory booth, an identical sensory booth complete with visual, auditory, and olfactory cues associated with cookie preparation (i.e. mixed immersion environment), and a full immersion testing room complete with same three context cues. We hypothesized that incorporating important contextual variables associated with cookie
preparation and consumption would improve the sensitivity, power and reliability of consumer hedonic data for cookie products consumed in a preparation setting as manifested by (1) a difference in hedonic data for products evaluated in immersive environments (full and mixed) compared to traditional sensory testing conditions, (2) more significant differences among products evaluated in immersive environments (full and mixed) compared to a traditional sensory testing environment and (3) more stable liking results in the full and mixed immersive environments compared to traditional testing booths when the same panelists repeat the test following ca. a 3-week hiatus, respectively. We also hypothesized that using immersive environments in hedonic testing of cookies would be more engaging than a traditional sensory testing environment as indicated by subjective assessments of the testing environments. We further hypothesized that the different levels of immersion utilized in the full immersion and mixed immersion environments during testing would differentially impact how much improvement was observed in the sensitivity, power and reliability of hedonic data and the subjective assessment of engagement when compared to results from the traditional sensory testing environment.
Chapter 3: Material & Methods

3.1 Participants

Fifty nine subjects (21 male and 38 female) ranging in age from 18 to 69 years old were recruited for this study using The Ohio State University Sensory Evaluation Center’s recruitment database. All of the 59 subjects recruited reported consuming cookies at least once a month and approximately half of them baked cookies at least once a month. Six participants failed to return for their second testing session so their results were excluded from the final data analysis. All participants were enrolled under informed consent (2013B0585) approved by The Ohio State University Institutional Review Board and a hard copy of the consent form was made available for participants. A copy of this can be reviewed in appendix A. Each subject participated in two experimental sessions that were scheduled approximately three weeks apart from each other to help minimize the likelihood that panelists would simply remember and copy prior responses. Each experimental session lasted approximately 30 minutes. At the conclusion of their second session participants received $20 in cash as incentive for their participation.

3.2 Stimuli

Four different brands of store bought chocolate chip cookies were used in this study: Chips Ahoy Chewy Chocolate Chip Cookies (Mondelēz International Group, East
Hanover, NJ), Kroger Bakery Chocolate Chip Cookies (Cincinnati, OH), Pepperidge Farm Soft Baked Montauk Chocolate Chip Cookies (Norwalk, CT) and Keebler Soft Batch Chocolate Chip Cookies (Kellogg’s, Battlecreek, MI). Cookie samples were purchased regularly at a local Kroger grocery store throughout the testing sessions and were stored in sealed containers at room temperature when not in use. All sample preparation took place within one day of the testing sessions to prevent samples from becoming stale. These four stimuli were selected based on the preliminary assessment that they differed enough in quality to obtain discriminating hedonic responses, were similar enough in appearance to prevent panelists from being able to memorize samples, and were all store bought chocolate chip style cookies. To further standardize their appearances a circular cookie cutter with a 1.5 inch diameter was used on all of the samples prior to their evaluation to give them a uniform size and shape. Each sample was packaged in a Dixie translucent plastic 2 oz. soufflé cup (Georgia-Pacific, Atlanta, GA) and covered with a Dixie translucent 2 oz. portion cup lid (Georgia-Pacific, Atlanta, GA) to prevent panelists from being able to see the appearance of the cookie until it was time for them to evaluate it. Each stimulus was assigned a 3-digit blinding code displayed by a sticker on the sample cup’s lid and new blinding codes were reassigned to each cookie for the second session. All four cookie samples were presented simultaneously in a row in a 3x4 Mirro aluminum muffin pan (Mirro Aluminum Company, Manitowoc, WI) at room temperature. The order that the samples were presented in was randomized among the panelists but remained the same for each individual across all three testing environments for both sessions to avoid changes in hedonic scores due to order effects.
(Mead & Gay 1995) and to help ensure that the main variable influencing the data were the environmental differences.

3.3 Procedure

At the beginning of each testing session a general instructions sheet was read to the panelists (see appendix B). It contained some general information such as what the panelist’s rights were as a research subject, the general testing procedure, and what was required of each panelist for them to earn their compensation. It also contained some more detailed information regarding proper sample evaluation methods and how to input answers using the tablet device provided. This general instructions sheet was made available in all three of the testing environments so that participants could refer back to it at any point of the testing session. After the general instructions sheet was read panelists were directed to their first testing environment. Once each panelist arrived in their first testing environment for their first testing session they began by signing an informed consent form and filling out a brief demographic questionnaire. This was the only time that this information was gathered so upon entrance to any subsequent testing environment panelists immediately began product evaluation. Panelists were then instructed to rate the acceptability of each of the four cookie samples presented to them on a 9-point hedonic scale. They were instructed to evaluate the samples in a serial monadic fashion, without re-tasting previous samples, taking a drink of water between each sample to cleanse their palate, and making sure to rate the correct cookie by ensuring that the number on the lid matched the number on the screen. While rating the
samples, panelists had access to their previous ratings for that testing environment but they did not have access to ratings given in previous testing environments or testing sessions.

Following sample evaluation, panelists completed a 19-item Engagement Questionnaire. The questionnaire was largely based on one used in a similar sensory study done by Bangcuyo et al., (2015) which was originally derived based on prior testing instruments developed by O’Brien and Toms (2010) to measure general engagement and by Witmer and Singer (1998) to specifically measure engagement within a virtual environment. The engagement questionnaire was composed of 19 questions designed to measure the level of agreement of consumers to specific statements and their responses were then sorted into eight dimensions to ultimately reflect a panelist’s engagement during the sample evaluation. The eight dimensions of engagement measured were: Usability, Environmental Aesthetics, Novelty, Involvement, Sensory Awareness, Immersion, Realism, and Distraction. The specific questions and statements that comprised each dimension are shown in appendix C. A minor change was made to the Engagement Questionnaire used in Bangcuyo et al., (2015) to derive these questions/statements. Within the Immersion dimension the previous used agreement statement of, “I lost track of time,” was replaced with, “I was absorbed in my testing environment,” to better reflect a panelists’ perceived level of immersion. Level of agreement to statements related to Usability, Environmental Aesthetics, Novelty, Involvement, and Immersion were collected using a 5-point Likert scale (coded -2 to +2) whereas responses related to the Sensory Awareness, Realism, and Distraction
dimensions were collected using a 7-point categorical scale ranging from 0 (none/not) to 6 (very). For each panelist, all of their dimensional responses were combined linearly to derive a singular measure of engagement referred to as the total engagement score which could theoretically range from -16 (distracting/not at all engaging) to +16 (extremely engaging).

In each of the two experimental sessions panelists were asked to complete the same sample evaluation and engagement questionnaire in all three of the environmental settings. All responses were recorded on a tablet device (8” Samsung Galaxy Tab 4, Suwon, South Korea) using Compusense At-Hand software (Ontario, Canada). Three panelists were recruited for each testing session although not every testing session contained three panelists due to scheduling issues. Each panelist started in one of the three testing environments and was then rotated to the other two testing environments as instructed by the two test supervisors upon completion of their sample evaluation/engagement questionnaire. Which environment the panelist started in and their order of rotation to the other two testing environments was randomized and counterbalanced across panelists and across the two different sessions. Before panelists entered their first testing environment, their cupped cookie samples were already placed in the metal muffin pan on top of the counter, along with the tablet in front of it, a filled water cup to the right, and the general instructions sheet on the left. This counter top layout was identical in all three of the testing environments. When rotating between environments panelists brought their water cup and assigned tablet device with them to the next environment. The test supervisors ensured that the panelist’s next samples were
ready upon entrance to their next testing environment so that they could immediately begin their evaluation. The only procedural difference between the three testing environments was that panelists were instructed to start the video component in the mixed immersion and full immersion environments before beginning their sample evaluation/engagement questionnaire. For all three testing environments, samples were evaluated under white light at an ambient temperature. Subjects returned after approximately three weeks and completed their second session of testing in an identical manner.

3.4 Testing environment - Traditional

Testing in the traditional environment took place in one of the two private sensory booths located in a laboratory room adjacent to the Immersive Technologies Laboratory at The Ohio State University. The sensory booth consisted of a 5’x7’ room with a counter top containing a spit sink and computer monitor with the only door leading to a small waiting area between the two sensory booths. Pictures of the three testing environments utilized can be seen in appendix D. Upon entrance panelists were seated at the counter, reminded of the instructions, and then began their sample evaluation and engagement questionnaire. The door to the booth was closed during their evaluation and only opened upon their completion.
3.5 Testing environment - Mixed immersion

Testing in the mixed immersion environment took place in the private sensory booth adjacent to the traditional sensory booth. The traditional and mixed immersion testing environments were separated by a 14’x8’ waiting area through which both booth entrances were located. The mixed immersion sensory booth was physically identical to that of the traditional testing booth. The only difference between the two testing environments was the addition of visual, audio, and olfactory context cues in the mixed immersion environment. The visual and audio context cues were provided by video footage of a man in the act of making cookies in a domestic kitchen setting. The video begins with the man combining common cookie ingredients in a tabletop mixer. As the video continues an oven timer goes off indicating a previous batch of cookies were finished so he pauses mixing to remove a tray of cookies from the nearby oven and places it on a cooling rack. The video ends with the man placing the newly created tray of cookies in the oven. The video was displayed to the panelist by a LG Flatron E2211 22” computer monitor (Seoul, South Korea) that was on the counter directly in front of the panelist. The audio from the video was relayed to the panelist using Sennheiser HD 280 Professional headphones (Wedemark, Germany). The video was recorded using a Nikon D5100 video camera (Tokyo, Japan) recording in 1080p HD at 30 frames per second with simultaneous recording of audio using a Tascam TM-2X stereo directional microphone (TEAC, Tokyo, Japan). The final video was edited to a length of 14.5 minutes to ensure that it would not end before panelists finished their evaluation as panelists were not forced to watch the entire video. In addition to the audiovisual stimulus a cookie aroma
(flavor code XL-879-653-9, Givaudan Flavors, Cincinnati, OH) was dispersed into the testing environment by bubbling air through the neat flavor at a constant rate to fill the room with a subtle but noticeable room aroma before panelists’ entry. Since the traditional and the mixed immersion testing environments shared a small waiting area outside of their entrances an Austin Air HealthMate HM400 air purifier (Buffalo, NY) was placed directly outside of the two testing environments and was constantly running during testing to ensure that the aroma being dispersed in the mixed immersion booth did not enter the waiting area or the traditional testing environment. Upon entrance panelists were seated at the counter, instructed to put on the headphones and start the video, and then begin their sample evaluation and engagement questionnaire. The door to the booth was closed during their evaluation and only opened upon completion to minimize possible odor contamination of the other two rooms. A testing supervisor was always present in the waiting space outside of the traditional and mixed immersion testing environment to answer any questions that panelists had, swap out trays of cookie samples between panelists, and to coordinate the proper rotation of panelists.

3.6 Testing environment - Full immersion

Testing in the full immersion environment took place in the Immersive Technologies Laboratory at The Ohio State University which is located directly adjacent to the room containing the traditional and mixed immersion environments. The Immersive Technologies Laboratory consists of one room with two sections separated by a door and a one-way mirror. The section behind the one-way mirror contained a counter
top containing controls over the visual and audio equipment present in the testing portion of the room. This location was where the other testing supervisor sat in order to control the video for this room and help with panelist rotation. The other portion of the room was an 8’x10’ testing area complete with a video wall composed of nine 48-inch high-definition LCD screens (model 47VS20-BAA, LG Electronics, Seoul, South Korea), surround sound tile speakers (Quam-Nichols, Chicago, IL) mounted in the ceiling and a tall chair and table placed directly in front of the video wall. The same video that was used in the mixed immersion testing environment was displayed on the video wall with the corresponding audio being played through the speaker system while the panelists completed their evaluation as in the mixed immersion environment. The same cookie aroma that was used in the mixed immersion testing environment was also dispersed into the full immersion testing environment in an identical fashion. Upon entrance panelists were seated at the table in front of the video wall in the testing portion of the room. The testing supervisor subsequently closed the dividing door between the two laboratory sections and started the video before panelists’ began their sample evaluation and engagement questionnaire.

3.7 Data Analysis

All data analyses were performed using SPSS 22 statistical software (IBM, Armonk, NY). Data transformations and calculation of the dimensional engagement scores, total engagement score, and change in mean liking scores between replicates for each cookie in each testing environment were accomplished using Excel (Microsoft
Corp., Redmond, WA). All data are reported as means with standard errors. Graphs shown were created using the GraphPad Prism 5 software (GraphPad Software Inc., La Jolla, CA).

3.7.1 Liking & Power

Differences in cookie liking under each condition (testing environment) were assessed using a 3-way, repeated measures Analysis of Variance (ANOVA); subject, cookie, and testing condition were main effects. With a significant condition*cookie interaction, we subsequently analyzed data from each condition separately using 2-way repeated measures ANOVA with post-hoc LSD and Tukey’s tests; subject and cookie were main effects under both analyses.

3.7.2 Reliability

The ability of cookie liking scores obtained in each testing condition to reliably predict cookie liking following a ca. 3-week hiatus was assessed based on the change in mean liking scores between replicates for each cookie in each condition. To do so we calculated the difference in each individual’s liking score between the two testing sessions for each cookie under each condition and then took the average for all of the participants. In addition, we combined the average change in mean liking scores across all cookies to further illustrate the differences in reliability between conditions. Further analyses on these data were conducted using a 3-way, repeated measures ANOVA (subject, cookie, and testing condition as main effects), with Tukey’s post-hoc analysis to identify the specific differences in reliability found between conditions.
3.7.3 Engagement

Likert data from the engagement questionnaire were coded -2 to 2 whereas categorical responses were coded 0 to 6. For each subject, coded data from relevant questions were averaged to generate dimensional scores for each testing condition. Dimensional scores were then linearly summed to generate the Total Engagement Score that represents a composite level of engagement. To determine whether dimensionality scores or the Total Engagement Score differed according to the testing environment, transformed data from each dimension were subjected to 2-way ANOVA with subject and testing condition as main effects. To determine whether the engagement disparity between the three testing environments was stable over time, a 3-way repeated measures ANOVA (subject, condition and replicate as main effects) was used to assess the average of the Total Engagement Scores across sessions for each condition in addition to the condition*replicate interaction term from the Total Engagement Scores. Additionally, when condition main effects were found to be significant under previous analyses, Tukey’s post-hoc analysis was utilized to identify the specific differences found between conditions.
Chapter 4: Results

4.1 Panelists’ cookie consumption/preparation habits

At the beginning of each panelist’s first testing session they reported their average cookie consumption and preparation habits as part of the demographic questionnaire. All of the 59 subjects originally recruited reported to consume cookies at least once a month with 26 panelists eating cookies an average of one to two times a month, 22 panelists eating cookies once to twice a week, 10 panelists eating them three to four times a week and one panelist claiming to eat them more than five times a week. Thirty-one panelists also claimed to bake cookies themselves less than once per month with 26 of them baking cookies one to two times a month and two participants claiming to bake cookies one to two times a week on average.

4.2 Liking

The use of immersive technologies to provide relevant contextual information during product evaluation affected the resulting consumer hedonic responses to the cookies. Results from the first session of testing for all panelists revealed a significant immersion level main effect ($F_{2,312}=6.535; p=0.002$) indicating that the average combined liking score for all cookies differed according to the environment in which they were evaluated. Post-hoc Tukey’s test showed that the full immersion testing environment had
the largest average liking score and was significantly different than the traditional testing environment and that the mixed immersion testing environment was not significantly different than the other two with an average liking score that fell almost directly in the middle of the other two environments.

A somewhat similar overall pattern for liking was seen during the second testing session when the same panelists evaluated the same four cookie samples following a ca. 3-week hiatus. Again a significant immersion level main effect was found ($F_{2,312}=5.837; p=0.003$) indicating that the average combined liking score for all cookies differed according to the environment in which they were evaluated. However, for this session, the post-hoc Tukey’s test showed that the mixed immersion environment had the highest average liking score and was significantly different than the traditional environment ($p=0.003$) while the full immersion environment had the second largest average liking score and was only found to be marginally significantly different than the traditional environment ($p=0.052$). The cookie by immersion level interaction was not significant for either of the testing sessions indicating that all of the cookies’ hedonic ratings were similarly impacted by the immersive conditions (Cookie*Immersion level interaction; 1st Session: $F_{6,312}=0.326, p=0.923$; 2nd Session: $F_{6,312}=0.813, p=0.561$). Both of these trends are clearly visible in figure 1A shown on the following page. Additionally, figure 1B displays the mean liking score given to each cookie in each of the testing environments for both sessions. Specific differences found in these data were discussed in the following section.
Figure 1. Cookie liking results obtained from panelists evaluating samples in each of the three conditions. A. Combined average hedonic ratings across cookies separated by condition and replicate. Letters above bars denote significant differences found between conditions using Tukey’s post-hoc analysis. B. Average hedonic ratings for each cookie separated by condition and replicate. Letters above bars denote significant differences found between conditions using LSD post-hoc analysis to further magnify differences found between cookies. Asterisks were used to denote marginally significant differences.
4.3 Power

The inclusion of relevant contextual information during product evaluation affected the power of the testing methodology with a more meaningful difference in power observed between the immersive and traditional environments during the second testing session. Tukey and LSD post-hoc analyses were used to identify all of the significant differences within the data as the difference in sensitivity between the two analyses allowed for two slightly different interpretations of the data, one being more conservative and the other more powerful respectively. Figure 1B on the previous page displays differences found using LSD post-hoc analysis. Additionally, the table in appendix F displays specific means and standard errors for the cookies in all of the environments across both sessions.

Tukey’s post-hoc analysis of the hedonic data for cookies that were evaluated in the first testing session showed that all three of the testing environments produced two homogenous subsets with two significant differences between cookie ratings found in the full and traditional environments and one significant and one marginally significant difference found in the mixed immersion environment. Using the more powerful LSD post-hoc analysis two significant differences in cookie liking were found in all three environments for the first testing session with a third marginally significant difference found in the full immersion environment. In either case, during the first replicate, the two post-hoc multiple comparison tests identified the fully immersive condition as being marginally more powerful than the mixed immersion or traditional conditions.
Tukey’s post-hoc analysis on the results from the second testing session did not find any significant differences between average liking scores for cookies evaluated in all three of the testing environments. However, one marginally significant difference was found between the Pepperidge Farm and Chips Ahoy brands of cookies (p=0.076) consumed in the mixed immersion testing environment suggesting that further trends may be identified using a more powerful post-hoc analysis. Using the more powerful LSD post-hoc analysis on the results from the second testing session revealed two homogenous subsets for the mixed immersion environment with two significant and one marginally significant differences found between cookie types. Under the same analysis for the second session, the full immersion environment had two marginally significant differences while the traditional environment still failed to produce a single difference between cookie ratings, significant or marginal. Whether this decrease in power for data collected in the traditional environment compared to the two immersive environments during the second testing session is due to the absence of contextual cues or a less engaging testing methodology is difficult to differentiate.

4.4 Reliability

To determine the reliability of hedonic data, the consumers re-evaluated the same cookies in all three of the testing environments following a ca. 3-week hiatus. In general, reliability between the two sessions was relatively high for all three of the testing environments which may have been due to obvious quality differences between the cookies used. For both testing sessions and across all of the environments the Chips Ahoy
cookie was consistently rated as the least desirable and the Kroger brand cookie was rated as the second least desirable. The Keebler and the Pepperidge Farm cookies were rated very similarly within each testing environment for both of the testing sessions—they even had the same exact average liking score for two different conditions—making comparisons based on the preference order of cookie samples difficult and less meaningful. Therefore, conclusions made about the reliability of the data collected in each environment were based on the magnitude of the differences found between each panelist’s hedonic scores for each cookie from the two testing sessions.

Figure 2 shown on the following page visually illustrates the differences found within each environment by graphing the average hedonic score for each cookie type for both of the testing sessions. Based on these line graphs the full immersion environment clearly appears to have produced the most reliable hedonic data between the two testing sessions. It also appears to show that the traditional testing environment produced more reliable average ratings for three out of the four cookies when compared to the mixed immersion environment’s line graph.
Figure 2. Average hedonic score for each cookie type across replicates separated by testing environment.
However, these line graphs are not the most effective way of illustrating the magnitude of this difference found between the two environments and do not reflect the variability associated with individual panelist’s hedonic ratings between the two sessions. Therefore we also calculated the difference in each individual’s rating between the two testing sessions for each cookie under each environment and then took the average for all of the participants. These results were graphed as the change in mean liking scores for each cookie in each testing environment as seen in figure 3 below. Based on this graph, the mixed immersion environment produced the largest difference in mean liking scores for three out of the four cookie samples with the traditional environment having the largest difference in mean liking score for one of the cookies and second largest difference for the other three. The full immersion environment had the smallest average change in hedonic score for all of the samples.

Figure 3. Change in mean liking scores for each cookie in each testing environment.
When the average change in mean liking scores was combined for all of the cookies, as can be seen in figure 4 below, we found that the mixed immersion environment had the largest difference between average testing session scores while the traditional environment had the second largest and the full immersion environment produced the smallest difference. However, the change in mean liking scores between testing environments were only fractions of a hedonic point away from each other with standard errors that were almost as large or larger than the actual change in mean liking scores. Consequently, there was only a marginally significant difference found for the immersion level effect of this data ($F_{2,312}=2.460, p=0.087$) with Tukey’s post-hoc analysis revealing this effect to be a result of a marginally significant difference found between the full and mixed immersion environments ($p=0.070$).

![Figure 4](image)

**Figure 4.** Average change in combined mean liking scores combined for all cookies in each testing environment. Letters above bars denote significant differences found between conditions using Tukey’s post-hoc analysis. Asterisks were used to denote marginally significant differences.
4.5 Engagement

In each testing environment for both of the testing sessions, subjects completed an Engagement Questionnaire that consisted of 19 items (for a list of the specific questions, see supplementary table in appendix B) comprising eight dimensions of engagement (Usability, Environmental Aesthetics, Novelty, Involvement, Sensory Awareness, Immersion, Realism, and Distraction). The use of immersive technologies to provide relevant contextual information during product evaluation had a significant effect on panelist engagement. A significant difference was found between testing environments for every dimension during at least one of the two testing sessions (shown by figure 5 on the page following this section and in detail by the table in appendix G). In particular, we found that the presence of relevant contextual information during product evaluation in the two immersive testing environments generally had a significantly positive impact on engagement data when compared to the traditional sensory testing environment. During both of the testing sessions panelists found the two immersive environments to be significantly more appealing (Environmental Aesthetics; 1st Session: $F_{2,104}=50.02$, $p<0.001$; 2nd Session: $F_{2,104}=85.50$, $p<0.001$); by engaging multiple senses (Sensory Awareness; 1st Session: $F_{2,104}=95.53$, $p<0.001$; 2nd Session: $F_{2,104}=126.60$, $p<0.001$); to increase enthusiasm for the test (Involvement; 1st Session: $F_{2,104}=22.12$, $p<0.001$; 2nd Session: $F_{2,104}=39.68$, $p<0.001$) and provide a testing environment that panelists felt was more ecologically valid than the traditional environment (Immersion; 1st Session: $F_{2,104}=64.11$, $p<0.001$; 2nd Session: $F_{2,104}=82.03$, $p<0.001$)
Three of the dimensions (Usability, Novelty, and Realism) did not produce any significant differences between environments during the first testing session but did so for the second session. Specifically, during the second session the panelists felt that both of the immersive environments were more consistent with their real-world experiences associated with cookie preparation and consumption compared to the traditional environment (Realism; $F_{2,104}=11.16$, $p<0.001$) and that the full immersion environment significantly assisted in their evaluations of the cookies (Usability; $p=0.043$) and enhanced curiosity associated with testing (Novelty; $p<0.001$) compared to the traditional environment with marginally significant results also produced between the mixed immersion and traditional environments for the same dimensions (Usability; $p=0.091$, Novelty; $p=0.068$). During the first session panelists found the immersive environments to be significantly more distracting compared to the traditional environment (Distraction; $F_{2,104}=5.23$, $p=0.007$) but only found them to be marginally more distracting during the second testing session (Distraction; $F_{2,104}=2.79$, $p=0.066$). The only dimension that found significant differences between all three of the testing environments was Immersion (1st Session: Traditional vs. Mixed or Full $p<0.001$, Mixed vs. Full $p=0.003$; 2nd Session: $p<0.001$ for all comparisons) suggesting that the level of immersion used might have produced varying degrees of ecological validity.

For each panelist, dimensional responses were combined linearly to derive a “Total Engagement Score,” ranging from -16 (distracting/not at all engaging) to +16 (extremely engaging). For both of the testing sessions, the Total Engagement Score only found significant differences between the traditional environment and the two immersive
environments (p<0.001 for both post-hoc comparisons) with no significant difference found between the mixed immersion environment and the full immersion environment suggesting that the application of immersive technologies improved overall engagement in both immersive environments. The Total Engagement Score data produced a marginally significant testing environment*testing session interaction (F_{2,260}=2.79, p=0.064) which would normally suggest that the level of engagement was not consistent between testing sessions for the three environments, but further investigation showed that this effect was a result of a significant decrease in the Total Engagement Score for the traditional testing environment from the first testing session to the second while the Total Engagement Score for the mixed and full immersion environments showed a slightly positive but not significant increase from the first session to the second. Evaluation of the average Total Engagement Scores across both testing sessions found each testing environment to be significantly different from each other (p<0.001) with the largest average score found in the full immersion environment (6.44+/-.61), followed closely by the mixed immersion environment (5.09+/-.59), and a negative average score for the traditional testing environment (-1.07+/-.68) suggesting that both of the immersive environments were significantly more engaging on average than the traditional environment with the full immersion environment found to be slightly more engaging than the mixed immersion environment.
Figure 5. Subjective assessment of panelist engagement in each of the three testing environments collected using the provided questionnaire. A. Average ratings for engagement dimensions for each testing environment separated by replicate. Letters above bars denote significant differences found between conditions using Tukey’s post-hoc analysis. Asterisks were used to denote marginally significant differences. B. Average total engagement score for each testing environment separated by replicate. Letters above bars denote significant differences found between conditions using Tukey’s post-hoc analysis.
Chapter 5: Discussion

The present study investigated the effect of utilizing immersive technologies to provide relevant contextual information within a sensory testing environment on the resulting hedonic data and panelist engagement. This was accomplished by comparing results in which subjects evaluated the same cookie samples in a traditional sensory testing booth, an identical sensory booth complete with visual, auditory, and olfactory cues associated with cookie preparation, and a full immersion testing room complete with same three context cues on two separate occasions. We hypothesized that improving the ecological validity of consumer hedonic testing by incorporating relevant visual, audio and olfactory cues through the use of immersive technologies would result in a more engaging testing experience and hedonic assessments that are more discriminating and reliable compared to those derived from traditional testing paradigms. We further hypothesized that the different levels of immersion utilized in the full immersion and mixed immersion environments during testing would differentially impact how much improvement was observed in the sensitivity, power and reliability of hedonic data and the subjective assessment of engagement when compared to results from the traditional sensory testing environment.

Indeed, we found the presence of contextual information and the degree of immersion employed to have an effect on the panelists’ evaluations of the cookie samples.
and their subjective assessment of engagement. Hedonic data collected in the two
immersive environments were found to be slightly more discriminating than data
collected in the traditional testing environment, particularly for the second session. The
two immersive environments also proved to be more engaging to consumer panelists
which may have contributed to the higher average hedonic ratings given in these
environments versus the traditional testing condition. In addition, the two immersive
environments remained equally engaging during both sessions whereas the traditional
testing environment was found to be less engaging during the second session. This may
have been partially responsible for the decrease in power observed from the first session
to the second in the traditional environment compared to the immersive environments.
Moreover, this finding suggests that the impact of immersive environments as designed
presently, was enduring, at least over the approximately 3-week test period evaluated
here. The reliability between the two sessions was relatively high for all three of the
testing environments but the mixed immersion environment produced the largest average
change in mean liking scores with the full immersion environment producing the smallest
average change. However, the changes in mean liking scores between testing
environments were only fractions of a hedonic point away from each other with standard
errors that were almost as large or larger than the changes themselves. This may have
been a consequence of limitations that are always associated with using untrained
consumer panelists in sensory testing such as memory effects and variable scale use
between panelists.
In general, the full immersion environment produced the best results out of the three environments. This finding along with the results from the Bangcuyo et al., (2015) study suggests that the level of immersion may play an important role in the utility of these methods. As such, dedication of an entire room to create a virtual reality testing environment might represent the most optimal design of such a testing practice. With that being said, the mixed immersion environment showed that it still might have the potential to be a viable alternative to a full immersion lab as it generally produced hedonic data of a higher quality than the traditional environment albeit not quite as compelling as the full immersion environment. Furthermore, the mixed immersion environment was found to be significantly more engaging than the same booth devoid of context and almost as engaging as a full immersion lab suggesting that the inclusion of relevant contextual cues may be more important than the scale of them. However, significant additional research is needed to fully understand the effect of using immersive technologies for consumer sensory testing and to identify the most optimal implementation of the methodology within the food industry.
Chapter 6: General Discussion

6.1 Immersion

New product development is necessary for most companies to stay competitive in the Food and Beverage Industry but approximately 80% of new food product launches fail to survive once in the marketplace (Redmond 1995). While multiple factors are likely responsible for this high failure rate, the inability of current consumer sensory testing methodologies to reliably measure consumers’ perceptions about a product and predict their resulting behavior is believed to be a major contributing factor (Deliza & MacFie 1996; Koster & Mojet 2007; Rosas-Nexticapa et al., 2005; Weiss et al., 2010). Despite numerous studies having found that extrinsic factors can completely change a consumers’ resulting hedonic assessment of a product, commonly used consumer sensory testing practices place panelists in environments intentionally designed to lack contextual information. A consequence of these sterile testing conditions and procedures that are often monotonous, is a general lack of consumer engagement during testing (Bangcuyo et al., 2015). This lack of engagement can result in uninformative hedonic data where panelists provide the same or similar ratings to all of the samples (Bangcuyo et al., 2015). Therefore, the lack of ecological validity and consumer engagement under the currently used, highly controlled, sensory testing methodology are thought to be two of the main
underlying reasons behind the inaccuracy of data collected under this approach (Bangcuyo et al., 2015).

Since virtual environments, by their very nature, are designed to be engaging and capable of providing extrinsic contextual information, they have the potential to be an ideal alternative to traditional sensory testing environments. A previous study by Bangcuyo et al., (2015) was one of the first to explore the utilization of immersive environments for consumer hedonic testing of food products. As in the present study, they found that the inclusion of relevant contextual cues within a sensory testing environment had an effect on the resulting hedonic data quality and consumers’ engagement level when compared to results obtained in a traditional sensory testing environment devoid of context (Bangcuyo et al., 2015). These results along with our study’s, suggest that incorporating immersive technologies in the field of consumer sensory testing has the potential to improve the ecological validity of testing conditions by restoring relevant contextual information resulting in data that can be more discriminating and reliable when compared to results obtained in a traditional sensory testing environment (Bangcuyo et al., 2015).

However, utilization of immersive technologies for consumer sensory testing using the testing design utilized in Bangcuyo et. al. (2015) would require companies to create or convert an existing room big enough to create a virtual lab in and would only allow them to test one consumer at a time. Since most companies who regularly perform consumer sensory tests use facilities capable of having small groups of consumers perform the test simultaneously, this one room setup would be far from ideal and the slow
through-put of the design might offset some of its potential cost savings. These two issues represent a significant limitation for adoption of this methodology by the food industry. Therefore one of the main objectives of the present study was to determine if implementation of the same immersive elements used in the virtual coffeehouse in a traditional sensory testing environment (i.e. the mixed immersion environment) could still positively impact the hedonic data and engagement of consumers as was seen by the virtual coffeehouse in the previous study. If proven effective this new methodology would allow companies the convenience of being able to easily modify their current sensory testing booths to contain visual, audio, and olfactory components relevant to the product being evaluated while maintaining their normal high throughput ability.

To determine the extent of the environmental factors’ effect on the resulting data we had panelists evaluate the same four cookies in all three of the testing environments for both sessions. Generally speaking, the mixed immersion environment produced better data than the traditional environment with results that were not quite as ideal as the full immersion environment. The data obtained from the engagement questionnaire found the two immersive environments to be significantly more engaging than the traditional environment with the mixed immersion environment only found to be slightly less engaging than the full immersion environment. This showed that the inclusion of relevant contextual information in a normal sensory testing booth made it significantly more engaging than the same sensory booth devoid of context and almost as engaging as the full immersion lab suggesting that the presence of context cues may be more important than the scale of them. However, the only real difference in the presentation style of the
contextual information between the two immersive environments was that the visual component was displayed on a much larger scale in the full immersion environment with the audio and olfactory components staying virtually the same in both environments. This suggests that engagement can be improved simply by changing the scale of the visual component in the testing environment although further research would be needed to validate that claim.

The average hedonic ratings for all cookies were larger in both of the immersive environments for both sessions compared to the traditional environment. Both of the immersive environments also maintained their ability to produce discriminating data for both sessions whereas the traditional environment only succeeded in doing so for the first session with no significant or marginally significant differences found between cookies during the second. Both of these trends may be a reflection of consumers staying more engaged in the cookie evaluation task when in the immersive environments compared to the traditional environment. Our data supported this in that the Total Engagement Score increased slightly across the two sessions for the two immersive conditions but actually decreased from the first session to the second in the traditional environment. This finding suggests that the impact of immersive environments as designed presently, was enduring, at least over the approximately 3-week test period evaluated here. Although further research is needed to fully evaluate this claim.

Reliability between the two sessions was relatively high for all three of the testing environments and the preference order of the cookies never changed significantly which may have been due to obvious quality differences between the cookies used. Evaluation
of the change in mean liking scores for each cookie across the two testing sessions found
the mixed immersion environment to have the largest average change and the full
immersion environment to have the smallest average change. However, the change in
average scores for cookies in all three of the testing environments were only fractions of
a hedonic point away from each other with standard errors that were almost as large or
larger than the actual change in mean liking scores making it difficult to make definitive
conclusions about the reliability of the data. This could have been a consequence of the
limitations that are always associated with using untrained consumer panelists in sensory
testing such as memory effects and variable scale use between panelists.

Therefore the present study showed that the use of immersive technologies to
include relevant contextual cues in a traditional sensory booth has the potential to
produce better hedonic data and keep panelists engaged when compared to traditional
sensory testing environments. However, since there are many factors that may have
affected our resulting data, further research is necessary to fully understand the most
optimal implementation of this practice.

6.2 Preparation Scenario

While the reasoning behind how the presence of contextual information may
affect consumer perception and their resulting behavior is likely complex and
multifaceted, one train of thought proposes that the context cues exert their influence
through the psychological construct of expectation (Deliza & MacFie 1996).
Expectations are important because they can strongly influence the perception of a
product before it is ever evaluated (Deliza & MacFie 1996). For instance, color cues associated with a beverage have been found to strongly influence the perceived intensity and quality of the product (Shankar, Simons, Shiv, McClure, Levitan, & Spence 2010; Morrot, Brochet, & Dubourdieu 2001). During food consumption, the degree of disparity found between previously formed expectations with the observed product attributes can affect the consumers’ resulting hedonic assessment and intent to purchase (Deliza & MacFie 1996). Historically, traditional sensory methodologies have viewed the expectations of sensory assessors as confounding (Deliza & MacFie 1996) and a potential source of bias (Cardello & Sawyer 1992). But this perspective comes from the same school of thought that views the inclusion of relevant context cues as confounding, and may be partially responsible for some of the shortcomings seen in current consumer sensory testing practices. How the use of relevant context cues can generate expectations and influence a consumer’s resulting opinions about a product are still not fully understood nor has it yet been proven if it is equally effective across all types of consumption scenarios or product types.

A previous study explored the effect of using a virtual coffeehouse in the sensory evaluation of five different coffee types (Bangcuyo et al., 2015). Interestingly, they found that when the five coffee samples were evaluated in the virtual coffeehouse the two most liked coffee samples were also the only two coffees that were obtained from actual coffeehouse purveyors (Bangcuyo et al., 2015). This was a consistent result for data obtained in the virtual coffeehouse but one that did not hold true for the data obtained in the traditional sensory testing environment (Bangcuyo et al., 2015). This finding suggests
that the presence of contextual information relevant to a coffeehouse generated specific expectations of a certain coffee profile that was confirmed and reinforced during the evaluation by the two coffee samples obtained from actual coffeehouses (Bangcuyo et al., 2015). Therefore, Bangcuyo et al., (2015) illustrated that inclusion of context cues relevant to the consumption of coffee at a coffeehouse was successful in generating expectations of products that would be found at an actual coffeehouse which ultimately resulted in better data quality compared to that obtained in the traditional testing environment devoid of context (Bangcuyo et al., 2015). This outcome showed the promising potential of using immersive technologies in consumer sensory testing of food products that are purchased and consumed at a retail location such as a coffeehouse. However, this retail based scenario only represents one of many ways that food products are regularly purchased and consumed with one of the more common methods of consumption being at the consumers’ own house following product purchase. Many of the food products regularly purchased at grocery stores or other outlets typically require some sort of preparation before it can be consumed. Therefore the act of preparation likely plays a role in the expectation and resulting opinions about a food product before it is able to be consumed.

In the present study we sought to build upon the investigation by Bangcuyo et al., (2015), by evaluating if the inclusion of context cues relevant to the preparation of food products in a home setting, in this case baking cookies in a kitchen, would have a similarly positive impact on the resulting hedonic data and engagement as was seen in the coffeehouse study. Based on our results both of the immersive environments that
contained relevant context cues about the preparation of cookies produced hedonic data of a better quality (as judged by discriminability and reliability measures) than did the traditional environment devoid of context. The panelists were also found to be more engaged in the two immersive environments that contained context cues which may have been partially responsible for the observed improvement of hedonic data quality. These findings showed that the inclusion of context cues relevant to the home preparation of food products in a sensory testing environment still had a positive impact on the resulting hedonic data like was seen by the retail scenario in the coffeehouse study. Therefore implementation of this consumer testing methodology or a similarly modeled one might also be beneficial to the large number of companies who produce food products meant to be prepared and consumed at home. However, since this category represents a very diverse and wide variety of food products further research would be necessary to confirm the effectiveness of this approach for other products since only cookie preparation and consumption were explored in the present study.

6.3 Limitations/Future Studies

The design used in this study represents one of many ways that immersive technology could be incorporated into the field of consumer sensory testing to provide relevant context cues to the participant during product evaluation. Due to the very limited amount of previous research exploring the concept, this study was intended to build upon the virtual coffeehouse study in an attempt to further understand the potential benefits and shortcomings associated with this testing methodology. In hindsight the study’s
design was not perfect and contained multiple elements that might be viewed as limitations and could have affected the resulting data. Future research can explore the effect of these limitations and develop new methodologies that limit their potential influence on the resulting data.

One of the potential limitations in the present study was the selection of the four cookies that were evaluated by the panelists. The video shown in the two immersive environments displayed cookies being made from scratch and freshly taken out of the oven which probably created an expectation of freshly made warm cookies. But the four cookies that the panelists actually evaluated were all store bought cookies served at room temperature. The decision to use these cookies helped reduce the variability between samples that would have occurred if panelists were given freshly made cookies served warm but as a consequence the products used presently might not have met the expectation created by the video shown. Additionally, each of the four cookies were rated similarly in all three of the environments for both sessions which might have been due to obvious quality differences being present between the samples. The combination of these two factors might have had a confounding effect on the quality of the resulting hedonic data potentially making it less effective at illustrating how differences between the three testing environments affected the results. Future studies on the topic should attempt to select products for evaluation that are as close as possible to the expectation being created by the context cues employed while still maintaining logistical testing feasibility.

Another part of the study design that might be viewed as a limitation was the fact that the manner in which the samples were presented did not change at all between the
In each of the environments for this study, cookie samples were packaged in translucent plastic 2 oz. soufflé cups with translucent lids displaying the blinding codes and were served in 4x3 muffin tins. Not changing the presentation style of the cookies between testing environments limited variability and helped to make the other contextual elements the only influencing factor between the three environments. However, we might have been able to further increase the ecological validity of the data obtained in the two immersive environments if we had chosen to make the presentation style of the cookies in those environments more like what might be found in a real-life scenario and less like how samples are typically presented under traditional sensory testing methodologies. Future studies could explore a variety of presentation style changes such as the use of real silverware or the absence of sample cups or blinding codes in an attempt to make the testing environment more consistent with real-life consumption circumstances.

In addition, the sample size used in this study might have been too small. The number of panelists recruited for the present study was based on the ability of the similarly designed study by Bangcuyo et al., (2015), to produce discriminating results with only 50 panelists. While the present study still produced statistically significant results, the testing methodology might not have been as powerful as the one utilized in the study by Bangcuyo et al., (2015). Therefore the use of a larger sample size might have further magnified the effect of differences between the three testing environments on the resulting data.
Another potential limitation of this study’s design that could have been approached in multiple ways was the means by which the video was utilized in the two immersive environments. One of the objectives of this study was to evaluate if the inclusion of context cues during consumer sensory testing would still be effective for food products that are prepared and consumed at home since the virtual coffeehouse study had already proven it to be effective for a product purchased and consumed at a retail location. Therefore the scenario revolved around a preparation based food product, in this case cookies, which is why it made sense to use a video that showed someone making cookies from scratch. However, we had panelists watch the video while they consumed and evaluated the cookie samples which meant that the task that they were performing was consumption based while the video and audio context cues shown to them were preparation based. While this method still had some congruency since both factors revolved around cookies it might have been more effective to have had the panelists watch the entire preparation based video before beginning their evaluation of the cookies. That way the video would have ended with the man pulling out a finished batch of fresh cookies right as the panelists received their cookie samples in order to maximize congruency between the two events. Ideally this approach could increase the ecological validity of the testing methodology and resulting data. Additionally, a consumption based video displaying a standard looking dinner table complete with a newly baked tray of cookies could have been used to follow up the preparation based video and shown during the time that panelists actually consumed and evaluated the cookies. This is just one of multiple ways that the testing procedure could have been
altered in an attempt to increase the ecological validity of the testing methodology but further research would be needed to fully evaluate the effect of such testing designs.

Additional future research that might provide further insight into the potential application of immersive environments within sensory testing environments would be to perform a similar study but this time excluding the use of a full immersion lab setup. The absence of this third environment might make it possible to more thoroughly identify differences in the quality of data obtained in a traditional sensory testing booth versus an identical booth complete with relevant context cues. Use of only these two testing environments compared to the three that were used in the present study could help reduce panelist fatigue and decrease the repetition of product evaluation making it harder to memorize samples and give repeat ratings. The results from the present study showed that the inclusion of relevant contextual information in a normal sensory testing booth made it significantly more engaging than the same sensory booth devoid of context and almost as engaging as the full immersion lab suggesting that the presence of context cues may be more important than the scale of them. Interestingly enough, the only real difference in the scale of the contextual effects between the two immersive environments was the visual component with the audio and olfactory components staying virtually the same. This suggests that engagement and the resulting hedonic data can potentially be improved simply by changing the scale of the visual component. Therefore future studies should explore using larger screens to convey the visual component in their sensory testing booth containing context cues for comparison with data obtained in an identical booth devoid of context.
Ultimately the approach of adding relevant context cues to previously existing sensory testing booths will make adoption by the food industry much easier and more ideal so future research should focus primarily on this approach. Furthermore, the use of immersive technologies in consumer testing of food products is novel and to date, there has been no previously published research delineating the approaches or obstacles associated with the implementation of such testing within the food industry. Therefore a significant amount of research is still needed to identify the parameters necessary for implementing immersive technologies in consumer sensory testing and to develop new methodologies that produce more reliable and discriminating hedonic data compared to traditional sensory testing paradigms.


Appendix A: IRB Subject Consent Form
The Ohio State University Consent to Participate in Research

Study Title: Using immersive technologies to study food perception and reward.

Researcher: Christopher T. Simons, Ph.D.

Sponsor: None

This is a consent form for research participation. It contains important information about this study and what to expect if you decide to participate.

Your participation is voluntary.

Please consider the information carefully. Feel free to ask questions before making your decision whether or not to participate. If you decide to participate, you will be asked to sign this form and will receive a copy of the form.

Purpose:
We are interested in the mechanisms underlying people’s perception, liking and intake of foods and beverages. In this particular study, we are investigating how contextual information influences these variables. When we consume foods and beverages, we are surrounded by external visual, auditory and aromatic cues. We believe this external information is processed along with the taste and smell of the foods to influence our perception, liking and intake of these products.

Procedures/Tasks:
You will be asked to evaluate a food or beverage product. In some cases, you may be asked to assess the sensory properties (taste, smell, texture or visual aspects) of these products and in other cases you may be asked to indicate how much you like or dislike each product. You may be asked to make your evaluations in a variety of testing environments including testing booths, our sensory lab, or even at a place of business or at a central location. After your evaluations, you may be asked to fill out a questionnaire assessing various aspects of your testing experience, eating habits or exposure to certain types of foods, restaurants and/or cooking experiences.

Duration:
Participation in this experiment will take no more than 30 min. In some cases, you may be asked to return to the laboratory at a subsequent time for further testing. If the study involves returning to the laboratory, the second experimental session will last no more than 30 minutes. Thus, your total time commitment will be no more than 60 minutes should you complete both
experimental sessions. In such instances that a second session is required, you will be notified prior to the onset of the first experimental session so you can decide if you want to participate.

You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you, and you will not lose any benefits to which you are otherwise entitled. Your decision will not affect your future relationship with The Ohio State University.

Risks and Benefits:
The food and flavor products that you will evaluate are comprised of ingredients that have been approved for use in foods by the United States Food and Drug Administration. However, if you have any known food allergies, please let us know immediately so we can confirm that the test product is safe for you to consume.

You will receive no direct benefit for participating in this study. However, the insight gained from your participation will give us a better idea of how various food attributes are processed by the brain to influence food perception and liking.

Confidentiality:
All information will be stored in a secure computerized database. At the onset of the experimental session, you will be asked to provide general demographic information including age, gender and ethnicity. In some cases additional information regarding eating and dietary habits may be obtained. These data will be collected using secured computerized data acquisition software or, on occasion, paper ballot. Data collected from paper ballots will be input into a secure computer at the earliest convenience and the paper ballot destroyed.

Efforts will be made to keep your study-related information confidential. However, there may be circumstances where this information must be released. For example, personal information regarding your participation in this study may be disclosed if required by state law. Also, your records may be reviewed by the following groups (as applicable to the research):

- Office for Human Research Protections or other federal, state, or international regulatory agencies;
- The Ohio State University Institutional Review Board or Office of Responsible Research Practices;
- The sponsor, if any, or agency (including the Food and Drug Administration for FDA-regulated research) supporting the study.

Incentives:
You will receive either course credit or a gift card in the amount of $20 per hour of time. At the conclusion of data collection, you can choose to be compensated with a gift card or course credit. In the event that you participate in an experiment that requires returning to the laboratory for multiple sessions, you will receive compensation at the end of each session.
Participant Rights:

You may refuse to participate in this study without penalty or loss of benefits to which you are otherwise entitled. If you are a student or employee at Ohio State, your decision will not affect your grades or employment status.

If you choose to participate in the study, you may discontinue participation at any time without penalty or loss of benefits. By signing this form, you do not give up any personal legal rights you may have as a participant in this study.

An Institutional Review Board responsible for human subjects research at The Ohio State University reviewed this research project and found it to be acceptable, according to applicable state and federal regulations and University policies designed to protect the rights and welfare of participants in research.

Contacts and Questions:
For questions, concerns, or complaints about the study, or if you feel you have been harmed as a result of study participation, you may contact the Principal Investigator, Christopher T. Simons at (614) 688-1489 or simons.103@osu.edu.

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.
Signing the consent form

I have read (or someone has read to me) this form and I am aware that I am being asked to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I voluntarily agree to participate in this study.

I am not giving up any legal rights by signing this form. I will be given a copy of this form.

Printed name of subject: ___________________________ Signature of subject: ___________________________

Date and time: __________ AM/PM

Printed name of person authorized to consent for subject (when applicable): ___________________________ Signature of person authorized to consent for subject (when applicable): ___________________________

Date and time: __________ AM/PM

Relationship to the subject: ___________________________

Investigator/Research Staff

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

Printed name of person obtaining consent: ___________________________ Signature of person obtaining consent: ___________________________

Date and time: __________ AM/PM
Appendix B: General Instructions Sheet

Today you will be evaluating a variety of cookies in three different rooms. Each of you will start in a different room and then rotate to the other two rooms as instructed by myself and my assistant. At the very start you will read an informed consent form that explains the general purpose of the research and your rights as a participant. At the end it will ask for your consent to participate as well as your first and last name. By consenting to participate in this study you are not only agreeing to participate today but also to return for an additional test that will be scheduled in the next few upcoming weeks. You will not receive your $20 cash compensation until the end of the second testing session. If you no longer want to participate in the study at any point then please notify myself or my assistant.

Following the consent form there’s a quick demographic survey and then the actual tests will begin. For each test you will be rating the acceptability of the cookies given to you followed by another small questionnaire. Please read the instructions on each screen before evaluating any samples. Once you have finished the test please notify my assistant or myself so that we can direct you to the next designated room.

When you evaluate the cookies, make sure that you are rating the correct one by ensuring that the number on the lid matches the number on the screen. They should be in the correct order, but it is always good to double check. Once you provide an acceptability rating for a cookie, do not go back and re-taste it. Finally, before tasting each sample, take a drink of water to rinse out your mouth. If you have any questions, during the test, just let us know.

All of your responses will be collected using a tablet device. When moving from one room to the next you will be taking your tablet with you along with your water cup. Please hold the tablet sideways while imputing answers. Some screens may require you to scroll down the page to view all of the questions so please be aware of that. When you get to your first testing room you will start by selecting the Demographic Questionnaire test. Once this is completed please notify my assistant or myself and we will instruct you on how to proceed from there.

These general instructions will be posted in each of the testing rooms in case you feel the need to refer back to them at any point. Otherwise enjoy the cookies!
### Appendix C: Engagement Questionnaire Organized by Dimension

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Statement/Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability</td>
<td>The testing environment assisted in my evaluations of the samples.</td>
</tr>
<tr>
<td>Environmental</td>
<td>The testing environment engaged my senses.</td>
</tr>
<tr>
<td>Aesthetics</td>
<td></td>
</tr>
<tr>
<td>Novelty</td>
<td>The testing environment incited my curiosity.</td>
</tr>
<tr>
<td></td>
<td>The testing environment distracted me.</td>
</tr>
<tr>
<td>Involvement</td>
<td>The testing environment was: (panelists ranked from very bored to very fun on a 5-point scale with neither boring or fun as the middle anchor)</td>
</tr>
<tr>
<td></td>
<td>I was engaged in the sensory task I performed.</td>
</tr>
<tr>
<td>Sensory Awareness</td>
<td>How completely were all of your senses engaged by the testing environment?</td>
</tr>
<tr>
<td></td>
<td>How much did the visual aspects of the testing environment involve you?</td>
</tr>
<tr>
<td></td>
<td>How much did the auditory aspects of the testing environment involve you?</td>
</tr>
<tr>
<td></td>
<td>How much did the olfactory aspects of the testing environment involve you?</td>
</tr>
<tr>
<td>Immersion</td>
<td>I was absorbed in my testing environment.</td>
</tr>
<tr>
<td></td>
<td>I felt like I was in a kitchen.</td>
</tr>
<tr>
<td>Realism</td>
<td>How disconnected did you feel from the testing environment?</td>
</tr>
<tr>
<td></td>
<td>How much did your experiences in the testing environment seem consistent with your real-world experiences?</td>
</tr>
<tr>
<td></td>
<td>How completely did you feel immersed in the testing environment?</td>
</tr>
<tr>
<td></td>
<td>How involved were you in the testing environment experience?</td>
</tr>
<tr>
<td>Distraction</td>
<td>How aware were you of events occurring in the real world around you?</td>
</tr>
<tr>
<td></td>
<td>How quickly did you adjust to the testing environment experience?</td>
</tr>
<tr>
<td></td>
<td>How much did the testing environment interfere or distract you from performing your sensory evaluation?</td>
</tr>
</tbody>
</table>
Appendix D: Testing Environment Pictures

A. Traditional sensory testing booth with counter top, spit sink, water cup, general instructions sheet, cookie samples presented in plastic sample cups in an aluminum muffin tin and a tablet device for recording responses. B. Mixed immersion testing booth in which contextual information was reintroduced into the testing environment. Visual information of a man making cookies was relayed using a computer monitor. Audio information was relayed using the pictured headphones. A subtle cookie aroma was dispersed into the testing environment by bubbling air through the neat flavor at a constant rate. C. Full immersion testing lab. Audiovisual information was relayed through multiple high-definition monitors. A subtle cookie aroma was dispersed into the testing environment by bubbling air through the neat flavor at a constant rate.
Appendix E: Diagram of Testing Environments

Figure 7. Diagram outlining the physical setup of the three testing environments.
Appendix F: Supplementary Table of Average Hedonic Ratings of Cookies

Table 1. Average hedonic scores for cookies separated by testing environment and replicate. Superscript letters denote significant differences found between cookies within a testing environment using LSD post-hoc analysis to further magnify differences found between cookies. Asterisks were used to denote marginally significant differences.

<table>
<thead>
<tr>
<th>Environment</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cookie</td>
<td>Mean+/-SE</td>
</tr>
<tr>
<td>Traditional</td>
<td>Chips Ahoy</td>
<td>5.06+/-0.24^A</td>
</tr>
<tr>
<td></td>
<td>Keebler</td>
<td>5.94+/-0.23^B</td>
</tr>
<tr>
<td></td>
<td>Kroger</td>
<td>5.40+/-0.23^AB</td>
</tr>
<tr>
<td></td>
<td>Pepperidge Farm</td>
<td>5.94+/-0.25^B</td>
</tr>
<tr>
<td>Mixed Immersion</td>
<td>Chips Ahoy</td>
<td>5.17+/-0.24^A</td>
</tr>
<tr>
<td></td>
<td>Keebler</td>
<td>6.15+/-0.23^B</td>
</tr>
<tr>
<td></td>
<td>Kroger</td>
<td>5.57+/-0.27^AB</td>
</tr>
<tr>
<td></td>
<td>Pepperidge Farm</td>
<td>5.94+/-0.26^B</td>
</tr>
<tr>
<td>Full Immersion</td>
<td>Chips Ahoy</td>
<td>5.26+/-0.25^A</td>
</tr>
<tr>
<td></td>
<td>Keebler</td>
<td>6.23+/-0.20^B</td>
</tr>
<tr>
<td></td>
<td>Kroger</td>
<td>5.81+/-0.27^B*</td>
</tr>
<tr>
<td></td>
<td>Pepperidge Farm</td>
<td>6.19+/-0.24^B</td>
</tr>
</tbody>
</table>
Appendix G: Supplementary Table of Average Engagement Scores

Table 2. Average engagement scores for each dimension separated by testing environment and replicate. Superscript letters denote significant differences found between dimensions within a testing environment using Tukey’s post-hoc analysis. Asterisks were used to denote marginally significant differences.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Environment Mean+/SE</td>
<td>Environment Mean+/SE</td>
</tr>
<tr>
<td>Usability</td>
<td>Traditional -0.09+/-.16&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional -0.21+/-.15&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Mixed Immersion 0.08+/-.13&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Mixed Immersion 0.17+/-.12&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 0.11+/-.14&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Full Immersion 0.23+/-.14&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td>Environmental</td>
<td>Traditional -0.38+/-.15&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional -0.81+/-.12&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>Mixed Immersion 0.94+/-.09&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Mixed Immersion 0.96+/-.10&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 1.00+/-.10&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Full Immersion 0.98+/-.12&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td>Novelty</td>
<td>Traditional 0.31+/-.09&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional 0.18+/-.07&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Mixed Immersion 0.36+/-.11&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Mixed Immersion 0.44+/-.09&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 0.46+/-.09&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Full Immersion 0.64+/-.10&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td>Involvement</td>
<td>Traditional 0.35+/-.10&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional 0.03+/-.09&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Mixed Immersion 0.85+/-.07&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Mixed Immersion 0.74+/-.08&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 0.97+/-.07&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Full Immersion 0.88+/-.08&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td>Immersion</td>
<td>Traditional -1.43+/-.12&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional -1.83+/-.05&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Mixed Immersion -0.23+/-.15&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Mixed Immersion -0.58+/-.13&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 0.30+/-.16&lt;sup&gt;C&lt;/sup&gt;</td>
<td>Full Immersion 0.17+/-.17&lt;sup&gt;C&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sensory Awareness</td>
<td>Traditional 1.58+/-.15&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional 1.08+/-.12&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Mixed Immersion 3.71+/-.15&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Mixed Immersion 3.77+/-.16&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 3.99+/-.16&lt;sup&gt;C&lt;/sup&gt;</td>
<td>Full Immersion 3.96+/-.17&lt;sup&gt;C&lt;/sup&gt;</td>
</tr>
<tr>
<td>Distraction</td>
<td>Traditional 1.62+/-.12&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional 1.62+/-.12&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Mixed Immersion 2.06+/-.14&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Mixed Immersion 1.94+/-.12&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 2.08+/-.13&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Full Immersion 1.93+/-.11&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>Realism</td>
<td>Traditional 1.28+/-.13&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional 1.06+/-.15&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Mixed Immersion 1.48+/-.09&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Mixed Immersion 1.50+/-.09&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 1.54+/-.08&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Full Immersion 1.65+/-.09&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td>Total Engagement</td>
<td>Traditional -0.01+/-.70&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Traditional -2.13+/-.60&lt;sup&gt;A&lt;/sup&gt;</td>
</tr>
<tr>
<td>Score</td>
<td>Mixed Immersion 5.13+/-.57&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Mixed Immersion 5.06+/-.60&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Full Immersion 6.30+/-.58&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Full Immersion 6.57+/-.62&lt;sup&gt;B&lt;/sup&gt;</td>
</tr>
</tbody>
</table>