The Pennsylvania State University

The Graduate School

College of Agricultural Sciences

EXPLORATION OF PRODUCT OPTIMIZATION USING CONSUMER-BASED TOOLS IN A COFFEE-FLAVORED DAIRY BEVERAGE

A Thesis in

Food Science

by

Bangde Li

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

May 2014

The thesis of Bangde Li was reviewed and approved* by the following:

John E. Hayes
Assistant Professor of Food Science
Thesis Adviser

Gregory R. Ziegler
Professor of Food Science

Robert F. Roberts
Professor of Food Science
Head of the Department of Food Science

*Signatures are on file in the Graduate School

ABSTRACT

Consumer insight plays a critical role in product development. A product can be optimized based on its formulations or sensory properties by maximizing consumer acceptability (i.e., liking). Both psychohedonic (sensation-liking) and physicohedonic (formulation-liking) models provide their own unique insights into consumer preferences. JAR scaling and ideal scaling have become quite popular to meet the demand of rapid optimization. In both these methods, a product attribute is measured for its dysfunctionality (delta) relative to one's ideal. Attribute delta (i.e., "Too Little" or "Too Much") estimates a subject's dissatisfaction (disliking) level with an attribute quality. Moreover, these methods differ in defining ideal levels on the scale. Dissatisfaction and liking may be two distinct constructs of consumer acceptability. We hypothesized that minimizing dissatisfaction and maximizing liking may yield different optimal formulations. The purpose of this study was to: 1) interpret consumer preference using physicohedonic and psychohedonic models; 2) investigate the difference between ideal scaling and JAR scaling for diagnosing attribute performance; 3) compare attribute delta (*Ideal_Delta* and *JAR_Delta*) models against liking models for product optimization of a coffee-flavored dairy beverage.

Coffee-flavored dairy beverages (n=20) were formulated using a fractional mixture design that constrained coffee extract, fluid milk (2% fat), sugar, and water. Participants (n=388) were randomly assigned into one of 3 research conditions that differed in ballot formats. Each participant tasted only 4 samples out of 20 using an incomplete block design. Samples were rated for liking and

intensities for four attributes--*sweetness*, *milk flavor*, *coffee flavor*, and *thickness*. Data were processed and treated differently to investigate specific research questions. Details are presented in the corresponding chapters.

The results show that: 1) both psychohedonic and physicohedonic models provide useful insights for product development; 2) ideal scaling and JAR scaling are very similar in estimating the attribute “Too Little” and “Too Much,” and these attribute deltas showed similar impacts on liking; 3) attribute delta and liking models yield different product optimization. That is what participants say they like differs from what they actually like.

Table of Contents

LIST OF TABLES viii
LIST OF FIGURES ix
ACKNOWLEDGEMENTS x

Chapter 1. Preface

Coffee-flavored dairy beverage 1
Consumer-based product optimization 3
Research objectives 6
Material and methods 7
Profiles of chapters 8

Chapter 2. Interpreting consumer preferences: physicohedonic and psychohedonic models yield different information about a coffee-flavored dairy beverage

Abstract 11
Introduction 13
Materials and methods 17
Results 23
Discussion 26
Conclusions 29
Acknowledgments 29

Chapter 3. Just-About-Right and ideal scaling provide similar insights into the influence of sensory attributes on liking

Abstract 30
Introduction 32
Materials and methods 34
Results 40

Discussion 47
Conclusions 52
Acknowledgments 53

Chapter 4. Product optimization: minimizing attributes "Too Little" or "Too Much" is not equivalent to maximizing overall liking

Abstract.. 54
Introduction 56
Materials and methods 60
Results.. 69
Discussion 79
Conclusions 87
Acknowledgments 89

Chapter 5. Conclusions and future work

Overall description 90
Conclusions 90
Suggestions for future works 93

Appendix A. Research method I

Product questionnaire.. 97
Demographic questionnaire 98

Appendix B. Research method II

Product questionnaire.. 99
Demographic questionnaire 100

Appendix C. Research method III

Product questionnaire.. 101
Demographic questionnaire 102

Appendix D. Validation liking study 103

Appendix E. Validation preference study 104

References 105

LIST OF TABLES

Table 2-1. Regularly consumed coffee-flavored products 18

Table 2-2. Sample formulations (in weight percentage) 20

Table 3-1. Frequency (%) of regularly consumed coffee-flavored beverages 36

Table 3-2. Sample formulations (in weight percentage) 37

Table 3-3. Effect of product and serving order on ideal ratings 42

Table 3-4. Effects of product and method on “Too Little” and “Too Much” 44

Table 4-1. Sample formulations (in weight percentage) in study I....................... 62

Table 4-2. Two optimal formulations (in weight percentage) in study II 63

Table 4-3. Frequency (%) of regularly consumed coffee-flavored beverages for participants in optimization study I 64

Table 4-4. Frequency (%) of regularly consumed coffee-flavored beverages for participants in validation study II 65

Table 4-5. *Overall_Liking* optimization model (n=388) 70

Table 4-6. *Liking_I* optimization model (n=127).. 72

Table 4-7. *Liking_II* and *Ideal_Delta* optimization model (n=129) 74

Table 4-8. *Liking_III and JAR_Delta* optimization model (n=132) 77

Table 4-9. Optimal formulations and predicted likings... 82

LIST OF FIGURES

Figure 1-1. Flow chart of prototypes formulation 7

Figure 2-1. Diagram of product optimization model 14

Figure 2-2. Effects graphs for psychophysical models. *Sweetness* (a,b,c), *coffee* (d,e,f) and *milk* (g,h,i) as a function of sucrose (a,d,g), coffee extract (b,e,h) and milk (c,f,i) concentrations 25

Figure 2-3. Effects graphs for *thickness* as a function of sucrose (a) and milk (b) concentrations and total solids content of the beverages (c) 26

Figure 3-1. Distribution of standard deviations of individual ideal ratings 41

Figure 3-2. Distributions of ideal intensity ratings using the mean ideal for each participant (n=114) 43

Figure 3-3. Attribute “Too Little” and “Too Much” comparison between two scaling methods 46

Figure 3-4. Influence of attribute “Too Little” and “Too Much” on *liking* 47

Figure 3-5. Asymmetrical impacts of *sweetness* and *coffee flavor* on *liking* 51

Figure 4-1. Contour plot for product optimization using *Overall_Liking* 71

Figure 4-2. Contour plots for product optimization using *Liking_I* (n=127) 73

Figure 4-3. Contour plots for product optimization using *Liking_II* and *ideal_Delta* (n=129) 75

Figure 4-4. Contour plots for product optimization using *Liking_III* and *JAR_Delta* (n=132) 78

ACKNOWLEDGEMENTS

To my committee: thank you for your guidance and questions that made my research more thorough. My future career will benefit from the experience of working with all of you.

To Dr. Hayes: thank you for bringing me into this program and helping to finish the current project research. You have been a great example of how to present your ideas to audiences effectively.

To Dr. Ziegler: thank you for your patience and intelligence. I have really enjoyed our conversations and discussions. I am amazed by your deep and wide knowledge in the field.

To Dr. Roberts: thank you so much for joining my committee. Your comments and questions have made my research so much more meaningful in the Food Science program.

To NIH: thank for the generous funding and support.

To my lab mates Alissa, Samantha, Nadia, Rachel A., Rachel P., Emma, Toral, Erin and Catherine: thank you all for helping conduct this study. You all have made my study and life at Penn State so productive.

To my family in China: thanks for your endless love, support, encouragement, and understanding. We missed you all so much. We look forward to seeing all of you in China soon.

To Sisson (Joy and Bill) family: thank you for the wonderful time you provided my family in this great country, and for helping us learn English and about American culture.

To my dear loves Jolynn (daughter), Bayllen (son), and my lovely wife Liying (Gillian): I couldn't have gotten through this program without your support. All of you are just perfect, and make sure I relax and have a fun time after school.

thank you!

Chapter 1

Preface

1. Coffee-flavored dairy beverage

Milk is a common drink in American families daily life. Between Oct 2012 and Oct 2013, milk production in the 23 major States has been increased about 1.2%, reaching a total production of 15.4 billion pounds (USDA, 2013). Milk and milk products are good sources of vitamin D, calcium, magnesium, and potassium (Ranganathan, Nicklas, Yang, & Berenson, 2005; Weinberg, Berner, & Groves, 2004). However, milk consumption among children and adolescents in the United States has declined since 1977-1978 (Hayden, Dong, & Carlson, 2013; Sebastian, Goldman, Enns, & LaComb, 2010). For most Americans, their consumptions of dairy products fall below the recommendations in *The Dietary Guidelines for Americans* (Hayden et al., 2013). Flavored milks are very popular among children and adults due to their desirable taste (Kim, Lopetcharat, & Drake, 2013). Flavored milks may provide a good opportunity to meet dietary guidelines for dairy products in the United States (Kim et al., 2013; Nicklas, O'Neil, & Fulgoni, 2013)

Frequently coffee is consumed after adding milk. Dairy-based coffee flavored beverages have become very popular in the past years (Boeneke, McGregor, & Aryana, 2007). There are conflicting views about the impact of adding milk to coffee. Antioxidant activity of espresso coffee dropped off when milk was added (Sánchez-González, Jiménez-Escrig, & Saura-Calixto, 2005). In contrast, another study showed that adding milk into a coffee beverage had a negligible effect on coffee antioxidant activities (Dupas, Marsset-Baglieri, Ordonaud, Ducept, & Maillard, 2006). However, no matter what the effect of milk on coffee might be in terms of antioxidants, consumers showed significant preference for milk-based coffee beverages over water-based coffee beverages (Cristovam et al., 2000). Dairy products have sensory properties, like mouth-feeling, oiliness, viscosity, sweetness and creaminess, which significantly influence consumer acceptability (Richardson-Harman et al., 2000). Milk can decrease the bitterness of coffee (Parat-Wilhelms et al., 2005). This might be due to mixture suppression in the brain (Lawless, 1979; Lawless, 1986), or be due to physio-chemical interactions (Bennett, Zhou, and Hayes 2012; Keast, 2008). Adding milk or cream alternatives into coffee has a significant impact on the coffee beverage's sensory properties, such as appearance, taste, and aroma (Richardson-Harman & Booth, 2006). Coffee with milk-added is perceived as ''sweet,'' ''creamy,'' and "milky," whereas water-based coffee is often perceived with either neutral or negative sensory perceptions, such as ''water-like,'' ''bitter,'' and ''bland'' (Petit & Sieffermann, 2007). Milk-added coffee not only can physically energize human body, but also can make a beverage milkier (Parat-Wilhelms et al., 2005). This energizing feature might be due to increased physiological arousal from caffeine. This may be why milk-based cappuccinos and lattes are so popular in the market.

Besides rich milk flavor, a dairy-based coffee beverage has a rich coffee flavor. Coffee flavor is generally regarded as a positive factor for consumer acceptance of a coffee beverage (Varela, Beltrán, & Fiszman, 2014). However, increasing coffee flavor by adding more coffee extract might also increase bitterness intensity. The bitterness in coffee beverage is generally regarded as a negative property (Cines & Rozin, 1982; Drewnowski, 2001). Coffee extract is a "complex" ingredient (Petit & Sieffermann, 2007). Therefore, a trade-off decision about the level of coffee extract has to be made to reach an optimal formulation. Insights gained from physicohedonic, psychophysical and psychohedonic models are helpful in making this decision.

2. Consumer-based product optimization

Optimization is commonly conducted using statistical models to maximize or minimize the corresponding variables that the developer is interested in (Gacula, 2008b). Traditionally, optimizaton is widely applied in engineering, such as by optimizing processing parameters (Ma et al., 2012). Consumers currently play a critical role in the process of product development (Costa & Jongen, 2006). Therefore, consumer insight is an important tool for product optimization (Chu & Resurreccion, 2004); Youn and Chung (2012) used consumer preferences to determine the optimal roasting temperature and time for a coffee-like beverage made from maize kernels. Dooley, Threlfall, and Meullenet (2012) optimized blended wines (Cabernet Sauvignon, Merlot and Zinfandel) by maximizing consumer acceptability (liking). However, the food industry shows increasing interest in rapid and easy tools for product optimization because time and cost are major concerns.

In the past decade, Just-About-Right (JAR) scaling has become popular in the food industry for product optimization (Popper & Gibes, 2004; Rothman & Parker, 2009; Xiong & Meullenet, 2006) because of its convenience and ease of use. Using JAR scaling, an attribute is evaluated for its appropriateness relative to an ideal level (Rothman & Parker, 2009; Worch, Dooley, Meullenet, & Punter, 2010). The ideal level is designated as “Just About Right” or “Just Right” in this method, and “Just About Right” or “Just Right” is fixed at the central point of scale. An attribute could be “Too Little,” “Too Much” or “Just About Right.” Particularly when an attribute is “Too Little” or “Too Much,” it can be optimized by increasing or decreasing attribute intensity by adjusting its corresponding ingredient concentration level. However, “Too Little” and “Too Much” qualities of an attribute do not always have equal influence on consumer acceptability (i.e., liking) (Xiong & Meullenet, 2006). JAR scaling is useful when a systematic solution (e.g., full formulation design) is not available because cost or time is a matter of concern. However, Stone and Sidel (2004) do not recommend replacing traditional experimental design with JAR scaling for product optimization. JAR scaling is criticized for its practice of combining the measurements of attribute intensity and consumer acceptability into the same scale (Moskowitz, Muñoz, & Gacula, 2008). This practice might create some biases (Rothman & Parker, 2009).

Alternatively, ideal scaling measures attribute perceived intensity and subjective ideal intensity separately (Gilbert, Young, Ball, & Murray, 1996; Rothman & Parker, 2009; van Trijp, Punter, Mickartz, & Kruithof, 2007; Worch, Le, Punter, & Pages, 2012a). Unlike JAR scaling, where the ideal level (i.e., "Just About Right" or "Just Right") is fixed at the central point of the scale, ideal scaling allows a participant to designate his/her hypothetical ideal level anywhere on the scale. Similarly, in ideal scaling, the attribute "Too Little" or "Too Much" refers to its perceived intensity below or above ideal intensity. The magnitudes (deltas) for attribute "Too Little" or "Too Much" can be estimated by the difference between perceived intensity and ideal intensity. However, comparisons of JAR scaling and ideal scaling for measurement of "Too Little" or "Too Much" are lacking in the literature. We hypothesized participants' ideal intensities differ from the central point of the scale, which might consequently influence the measurement of "Too Little" and "Too Much," and distort their influence on liking.

Notably, both ideal scaling and JAR scaling measure an attribute for its level of dysfunction, i.e., how far an attribute's perceived intensity deviates from one's ideal level. This level of dysfunction is estimated by the difference between perceived intensity and one's ideal level. Presumably, consumers would dislike or be dissatisfied with a product with a dysfunctional attribute. So to some extent, ideal scaling and JAR scaling measure a subject's dissatisfaction level for an attribute's quality. The more an attribute deviates from one's ideal level, the more dysfunctional an attribute would be. Correspondingly, consumer dissatisfaction for that attribute presumably increases as the deviation from one's ideal level increases. Critically, attribute dissatisfaction and liking might be two different constructs for measuring consumer acceptability. As a result, the factors driving liking and dissatisfaction might differ. In the Kano model, consumer dissatisfaction is not simply the opposite of satisfaction (Berger et al., 1993; Kano, Seraku, Takahashi, & Tsuji, 1984). Driving factors for satisfaction are “satisfier” and “performance” attributes. In contrast, factors driving dissatisfaction are “must-be” and “performance” attributes (Bi, 2012; Li, 2011). Additionally, prior works shows optimal formulations achieved by JAR scaling differ from those predicted by hedonic scores (Epler, Chambers IV, & Kemp, 1998; Shepherd, Smith, & Farleigh, 1989; Vickers, 1988). Herein we hypothesized that attribute delta in ideal scaling or JAR scaling differed from the measurement of liking in terms of product quality, and as a result, optimal formulations would differ when a) minimizing attribute delta or b) maximizing consumer liking.

3. Research objectives

The initial goal for this project is to optimize a coffee-flavored dairy beverage to extend current chocolate-flavored milk sold by the Creamery facility at the Pennsylvania State University. Due to our lean experimental design and rich dataset, we explored several interesting topics for product optimization. Specific purposes of this study include: 1) interpretation of consumer preference using psychophysical and psychohedonic models. 2) investigation of the difference between ideal scaling and JAR scaling for diagnosing attribute performance. 3) comparison of attribute delta (*Ideal_Delta* and *JAR_Delta*) model with a liking model for product optimization.

4. Material and methods

Prototypes (n=20) were formulated using a mixture design with constrained variables of coffee extract (3.0-5.0 wt %; Autocrat Sumatra 1397, Autocrat Natural Ingredients, Lincoln, RI), sucrose (5.0-8.0 wt %), milk (35-55 wt %, 2% fat), and water (35-55 wt %). These components accounted for 99.8% of the individual formulations. A constant amount of pectin (0.2 wt %; Grinsted® SY, Dupont Danisco) was added to all the samples. For convenience, pectin was mixed with sucrose completely before blending them with water, milk, and coffee extract to make sample batches (Figure 1-1). Batches were heated up to 72º C and held 15 seconds. After the heat treatment, prototypes were removed and rapidly transferred into sanitized carboys. Prototypes were cooled quickly by storing them in a fridge (~4.5ºC), and stood for about 24 hours before serving.

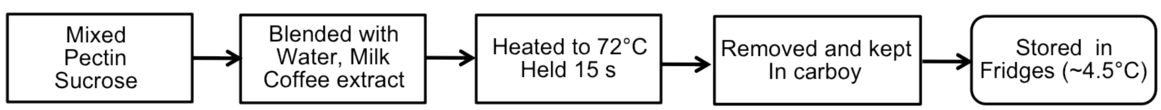

Figure 1-1. Flow chart of prototypes formulation

The project was interested in evaluating product optimization using consumer liking, ideal scaling and JAR scaling. For that purpose, the consumer study was designed using three research methods that differed in the questionnaires that were used (see Appendix A, B, and C). Each participant tasted only 4 samples out of 20 within one of the research methods, i.e., an incomplete random block design was applied. For all three methods, overall liking for samples was measured. Besides overall liking, each sample was also diagnosed for four attributes (*sweetness*, *milk flavor*, *coffee flavor*, and *thickness*). Data were processed and investigated for specific research questions.

5. Profiles of chapters

In chapter 2, both psychohedonic model and physicohedonic models were constructed to investigate the effects of formulation variables (sucrose, milk, coffee extract) and sensory properties (*sweetness*, *milk flavor, and coffee flavor*) on liking using simple linear regression models. The psychohedonic model showed a better prediction for liking than did the physicohedonic model. In the physicohedonic model, coffee extract showed a negative impact on liking. In contrast, coffee flavor was a positive factor to liking in the psychohedonic model. Coffee extract is a perceptually complex ingredient. Intensifying coffee flavor by adding more coffee extract might increase the bitterness of the beverage. Thus it is meaningful to consider both models simultaneously during product optimization.

In chapter 3, a comparison between ideal scaling and JAR scaling was conducted to investigate how setting ideal levels in the two methods potentially affects attribute “Too Little” and “Too Much” measurements and their influence on liking. The comparison showed no difference between the two methods in measuring the attribute “Too Little” and “Too Much.” Both ideal scaling and JAR scaling identified *sweetness* and *coffee flavor* as critical factors driving consumer liking. In multiple linear regression, JAR scaling explained more variance in consumer liking than ideal scaling did.

In chapter 4, liking and dissatisfaction (attribute deltas) models were addressed and compared for product optimization. Attribute delta was defined as the deviation between perceived intensity and one's ideal intensity, reflecting a consumer's dissatisfaction level toward attribute performance. We believe dissatisfaction is a different construct from liking for measuring consumer acceptability. The optimal formulation for a coffee-flavored dairy beverage obtained by minimizing *Ideal_Delta* or *JAR_Delta* (disliking model) had more coffee extract and less milk and sucrose than the optimal formula obtained by maximizing *liking*. Participants generally liked weaker, milkier and sweeter coffee more than that suggested by their ideal scale and JAR ratings. In a head to head validation study of the two optimal formulas, participants preferred the sample formulated using the liking model. This is consistent with the idea that consumers do not know what they want. Thus, asking consumer panel to design a product via JAR scaling or ideal scaling may be misleading. Instead, maximizing overall liking is a superior tool for product development. In addition, compared to ideal scaling, JAR scaling is more sensitive to the changes of formulation variables.

In chapter 5, conclusions and some suggestions for future work are summarized. Main conclusions include: 1) what consumers desire to have and what they say they would like are two distinct constructs; 2) JAR scaling did a better job than ideal scaling did in product optimization, and is recommended for the food industry in terms of ease and convenience; 3) both psychohedonic and physicohedonic models offer deep insights into understanding consumer preference. Main suggestions for future work include: 1) the range of coffee extract concentration should be expanded to reach an appropriate optimal formulation; 2) Coffee extract will generate not only coffee flavor but also other sensory properties that might have a critical influence on liking and preference, such as bitterness, color; the influence of these attributes on quality should be diagnosed and compared to further understand product overall performance and consumer liking and preference behavior in the validation study.

Chapter 2

Interpreting consumer preferences: physicohedonic and psychohedonic models yield different information in a coffee-flavored dairy beverage

Accepted by Food Quality and Preference

Abstract

Designed experiments provide product developers feedback on the relationship between formulation and consumer acceptability. While actionable, this approach typically assumes a simple psychophysical relationship between ingredient concentration and perceived intensity. This assumption may not be valid, especially in cases where perceptual interactions occur. Additional information can be gained by considering the liking-intensity function, as single ingredients can influence more than one perceptual attribute. Here, 20 coffee-flavored dairy beverages were formulated using a fractional mixture design that varied the amount of coffee extract, fluid milk, sucrose, and water. Overall liking (*liking*) was assessed by 388 consumers using an incomplete block design (4 out of 20 prototypes) to limit fatigue; all participants also rated the samples for intensity of coffee flavor (*coffee)*, milk flavor (*milk)*, sweetness (*sweetness)* and thickness (*thickness)*. Across product means, the concentration variables explained 52% of the variance in *liking* in main effects multiple regression. The amount of sucrose (β = 0.46) and milk (β = 0.46) contributed significantly to the model (p's <0.02) while coffee extract (β = - 0.17; p = 0.35) did not. A comparable model based on the perceived intensity explained 63% of the variance in mean

liking; *sweetness* (β = 0.53) and *milk* (β = 0.69) contributed significantly to the model (p's <0.04), while the influence of *coffee* flavor (β = 0.48) was positive but marginally ($p = 0.09$). Since a strong linear relationship existed between coffee extract concentration and coffee flavor, this discrepancy between the two models was unexpected, and probably indicates that adding more coffee extract also adds a negative attribute, e.g. too much bitterness. In summary, modeling liking as a function of both perceived intensity and physical concentration provides a richer interpretation of consumer data.

1. **Introduction**

Optimization is an efficient and practical tool for product developers (Ares, Varela, Rado, & Gimenez, 2011; Dutcosky, Grossmann, Silva, & Welsch, 2006) to achieve a competitive product in the market (Stone & Sidel, 2004; Villegas, Tarrega, Carbonell, & Costell, 2010). Not only can an optimization technique define an optimal product (Dutcosky et al., 2006), but also help evaluate effects of independent variables on the response variables. Traditionally optimization techniques have been widely used in engineering. For example, response surface methodology (RSM) has been used to explore the optimal roasting temperature and time in terms of yield, levels of free sugar, phenolic compounds, antioxidant activity, and sensory preference for a coffee-like beverage from maize kernels (Youn & Chung, 2012). In the current marketplace consumers are more influential in the product value chain and play an important role in the process of new product development (Costa & Jongen, 2006). Thus, it is important to integrate consumer insights into each step of product development (Brunso & Grunert, 2007).

Product sensory properties directly influence consumer preferences and purchases (Mitchell, Brunton, & Wilkinson, 2009). The concepts and techniques of optimization, such as response surface methodology (RSM) (Modha & Pal, 2011), Euclidian distance ideal point mapping (EDIPM) (Meullenet, Lovely, Threlfall, Morris, & Striegler, 2008), preference mapping techniques (Greenhoff & MacFie, 1999), and Landscape Segment Analysis (LSA®, IFPrograms), have been applied in sensory science to explore consumer-defined optimal product characteristics. Here we employed a fractional, constrained mixture design for formulation and an incomplete block design for sensory analysis using untrained consumers.

Three distinct models are useful to properly integrate consumer insights into product development: physicohedonic (concentration-liking) models, psychophysical (concentration-sensation) models, and psychohedonic (sensation-liking) models (Figure 2-1). Each model provides unique insights and meaningful feedback for product development. Physicohedonic and psychohedonic models are of more interest to product developers due to their ability to offer directional solutions to questions of formulation, while psychophysical models offer insights into the relationship of physicohedonic and psychohedonic models.

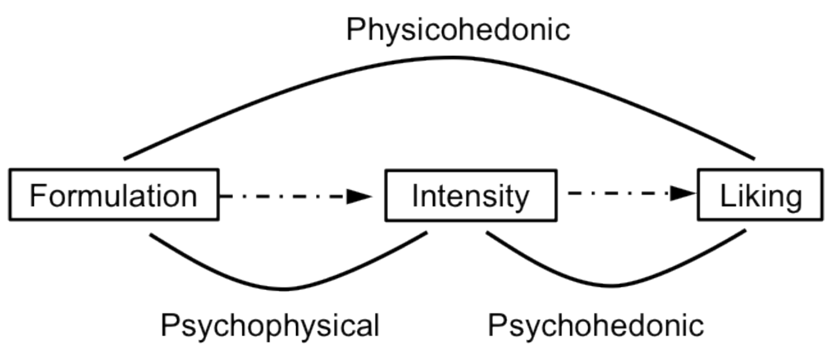

Figure 2-1. Diagram of product optimization model

Physicohedonic models are based on design variables (i.e., formulation) and consumer acceptability (i.e., liking). For example, consumer liking was modeled as a function of formulation to identify an optimal blended wine using a mixture design (Dooley, Threlfall, & Meullenet, 2012). Using this approach, the influence of design variables on response variables can be investigated, and optimal products can be described in terms of design variables. From the product developer's perspective, the physicohedonic models provide a practical and actionable solution; helping to determine which design variables are critical and how the product can be improved. However, this approach may be an oversimplification, as it assumes a simple relationship between concentration and sensation. Critically, this assumption may not be valid (Keast & Hayes, 2011), especially in cases where perceptual interactions may occur, when individuals perceive similar intensities from different physical concentrations, or when a single ingredient may contribute more than one sensory attribute. For example, adding more coffee extract into a coffee-flavored beverage might increase coffee flavor, which is assumed to be a positive factor for consumer acceptance, but may also increase bitterness that may be detrimental to consumer liking (Moskowitz & Gofman, 2007). The third case is seen for many non-nutritive sweeteners; in the asymptotic portion of the sweetness dose response curve, adding more acesulfame-K does not increase sweetness, but does increase bitterness (Schiffman, Booth, Losee, Pecore, & Warwick, 1995). In addition to changes in the intensity of a sensory response, the nature of the sensory property might be perceived differently with increase in concentration. For example, at a low titanium dioxide level, cheese looked opaque, but turned too white when extra titanium dioxide was added (Wadhwani & McMahon, 2012). The relationship between concentration and attribute sensory intensity is not normally linear (Hough, Sanchez, Barbieri, & Martinez, 1997).

In contrast, psychohedonic models link consumer acceptability to the perception of a product's sensory attributes (Greenhoff & MacFie, 1999; Meullenet et al., 2008). Fundamentally, psychohedonic models are meaningful and important (Keast & Hayes, 2011), because they can give direct feedback about factors driving consumer acceptance based on their sensory impact (Lovely & Meullenet, 2009). However, psychohedonic models may be less actionable. First of all, interactions between sensory properties are common (Hayes & Duffy, 2007;Wadhwani & McMahon, 2012; Xiong & Meullenet, 2006). Consequently, changes in one attribute might influence the perception of other properties (Hayes, Sullivan, & Duffy, 2010). Second, whatever findings are achieved from a psychohedonic model, further action on the product would typically be carried out by altering the formulation. Additionally, a sensory perception might be a function of multiple chemical components or design variables, e.g. adding either more milk or sucrose into a dairy-based beverage increases its thickness. As a result, the psychohedonic model may not directly indicate a workable solution for product improvement. Given that sensation (perceived intensity) is an intermediate variable between formulation and liking, we would expect sensation to be a better predictor of liking than concentration; indeed, perceived sweetness and creaminess are better predictors of liking than fat and sucrose concentration (Hayes & Duffy, 2008).

The objective of creating either physicohedonic or psychohedonic models is an understanding of consumer needs. Psychophysical models are useful for understanding and explaining conflicting information from physicohedonic and psychohedonic models during product development. To increase the likelihood of creating a successful product, it may be advantageous to study a two-stage concentration-sensation-liking model in addition to the simpler concentration-liking model. The present study was originally designed to optimize formulation of a new ready to drink beverage (a coffee-flavored milk) for retail sale in a campus facility. Here, we explore the insight gained from moving beyond a physicohedonic model to a multipart model that considers psychophysical and psychohedonic relationships separately.

2. Materials and methods

Participants were randomly assigned to one of three research conditions (described below). All the participants rated overall liking (*liking)* as well as the intensity of sweetness (*sweetness*), coffee flavor (*coffee*), milk flavor (*milk*), and thickness (*thickness*).

2.1 Ethics statement

Procedures were exempted from IRB review by the Penn State Office of Research Protections staff under the wholesome foods exemption in 45 CFR 46.101(b)(6). Participants provided informed consent and were compensated for their time.

2.2 Subjects

A total of 388 participants (110 men) were recruited ahead of time using an existing participant database maintained by the Sensory Evaluation Center at Penn State, or via staff intercepts in public spaces in and around the Food Science Department at Penn State.

To qualify for participation, individuals had to indicate they drank coffee or coffee-flavored beverages regularly (Table 2-1), and did not have any food allergies. About 40% of the consumers (n=155) were between 18-27 years old, 72 were 28-37, 56 were 38-47, 75 were 48-57, 26 were 58-67, and only 4 were over 67 years old. The majority (~77%) were White (n=298), while 59 identified themselves as Asian or Pacific Islanders, 9 as African or African American, and 11 did not report a race.

Table 2-1. Regularly consumed coffee-flavored products

Products	Frequency (%)
Cappuccino	23.7
Latte	31.4
Black coffee	25.0
Iced coffee	32.4
Coffee with milk, cream, and/or sucrose	60.0

Note: This is a “check all that apply” question, so the sums in a column may exceed 100%.

2.3 Sample formulation and preparation

Twenty coffee-flavored dairy beverages were formulated using a fractional mixture design with four constrained variables: coffee extract (3.0-5.0 wt %; Autocrat Sumatra 1397, Autocrat Natural Ingredients, Lincoln, RI), sucrose (5.0-8.0 wt %), milk (35-55 wt %, 2% fat, Berkey Creamery, University Park, PA), and water (35-55 wt %). These components accounted for 99.8% of the individual formulations. A constant amount of pectin (0.2 wt %; Grinsted® SY, Dupont Danisco) was added to all the samples. The exact composition of each formula is shown in Table 2-2. Pectin solutions were first prepared by blending pectin into the water. Coffee extract, milk, and sucrose were added to pectin solutions.

Table 2-2. Sample formulations (in weight percentage)

Product[1]	Milk (%)	Water (%)	Coffee extract (%)	Sucrose (%)	Solid content (%)[2]
1	35.93	54.89	3.99	4.99	10.02
2	45.24	45.24	4.32	4.99	11.11
3, 19	36.93	54.89	2.99	4.99	9.90
4, 9	53.89	34.93	2.99	7.98	14.75
5	34.93	51.90	4.99	7.98	13.14
6, 18	34.93	54.89	4.99	4.99	10.15
7	44.91	43.41	4.99	6.49	12.73
8	35.93	54.39	2.99	6.49	11.29
10	54.89	34.93	4.99	4.99	12.32
11	44.41	44.41	2.99	7.98	13.71
12, 16	54.39	34.93	3.99	6.49	13.53
13	54.89	36.93	2.99	4.99	11.86
14, 17	34.93	53.89	2.99	7.98	12.68
15	51.90	34.93	4.99	7.98	14.99
20	34.93	52.89	3.99	7.98	12.91

[1]Samples in the same row share the same formulation.

[2]Calculated from the solids content of the ingredients.

Batches were heated to 72 °C to assure that the sucrose was completely dissolved, the pectin dispersed, and the product was safe for consumption. The finished samples were kept at refrigeration temperature (~4°C) for at least 24 hours before serving. Two ounces of the coffee milk were served in 4-oz Solo transparent plastic cups (Solo Cup Company, Urbana, IL).

2.4 Product testing

Data were collected using Compusense *five*® (Compusense Inc., Guelph, ON, Canada) software. Participants were randomized to 1 of 3 test conditions upon entering test booths. In method I (n=127), only *liking* and attribute intensities were collected. In method II (n=129), participants rated *liking*, attribute intensities, and their ideal attribute intensities on separate, appropriately-worded line scales. In method III (n=132), *liking* was collected, and attribute appropriateness was assessed with Just-About-Right (JAR) scales. The ideal intensity and JAR data were not used here and will be reported elsewhere.

Liking was assessed using a standard 9-point hedonic scale (1 = "Dislike Extremely", 5 = "Neither Like Nor Dislike", and 9 = "Like Extremely") (Peryam & Pilgrim, 1957). Attribute intensities, both perceived and ideal, were measured using continuous line scales (0-100); two descriptive anchors were placed on 10% and 90% of these scales, representing low intensity (e.g., "Not At All Sweet") and high intensity (e.g., "Extremely Sweet"). Just-About-Right (JAR) scales were designed as continuous line scales with three descriptive anchors, low intensity (i.e., "Much Too Weak") on the left end, "Just About Right" at the middle, and high intensity (i.e., "Much Too Strong") on the right end.

Demographics and consumption behavior for coffee-based beverages were collected at the end of the session, after all sample evaluations.

To minimize fatigue, participants received 4 formulas out of 20 in an incomplete block design. The samples were served in a monadic sequential order, with a two-minute mandatory break between samples. During the break, participants were asked to rinse with room temperature (22°C) filtered water to reduce potential carry-over effects.

2.5 Statistical analyses

Data were analyzed using JMP® version 9.02 (SAS Institute Inc.). Analysis of variance (ANOVA) was conducted to detect effects of test conditions (method), product, and their interaction on *liking*. In the ANOVA model, panelist was a random variable nested within the method factor; method, product and their interaction were treated as fixed effects. Similar to multiple linear optimization models reported in the field (Johnson & Vickers, 1988; Schutz, 1983; Stone & Sidel, 2004), two linear regression models were fitted to diagnose and compare effects of formulation variables (sucrose, milk and coffee extract) and perceived attribute intensities on *liking*, i.e., a physicohedonic (formulation-liking) model and psychohedonic (intensity-liking) model. In these two models, means of liking and intensity data were regressed using JMP®. Similarly, attribute intensities were regressed on formulation variables using multiple linear regression in JMP®.

3. Results

3.1 Influence of research method on *liking*

To justify aggregation of the data, the effect of research method on *liking* was determined. In the ANOVA model, 52% of the variance in *liking* was explained by product (i.e. formulation), method, and participant. As expected, *liking* differed as a function of product ($F_{19,1300} = 8.66$, $p<0.0001$). The effect of method on *liking* was not significant ($F_{2,374.5} = 0.75$, $p=0.47$), nor was the product by test method interaction ($F_{38,1297} = 1.33$, $p=0.09$), indicating there was no systematic difference in *liking* resulting from the test methods. Therefore, *liking* data were combined across methods for the remaining analyses.

3.2 Effect of formulation on *liking*

In the physicohedonic model, concentration variables (amount of coffee extract, milk, and sucrose) explained 52% of the variance in *liking* in main effects multiple regression (fitted model: *liking* = 4.0+2.8**milk*-10.3**coffee*+17.6**sucrose*, p=0.008). The amount of sucrose (β= 0.46) and milk (β= 0.46) contributed significantly to the model (p's <0.02) while coffee extract (β = -0.17) did not (p = 0.35). The amount of sucrose and milk were equally important to *liking* in this model. Although not significant, greater amounts of coffee extract seemed to negatively influence *liking*.

3.3 Relationship between formulation and perceived intensity

Attribute intensities (*sweetness*, *milk*, *coffee* and *thickness*) were regressed on the concentrations of formulation variables (i.e., sucrose, milk, coffee extract and total solids) using multiple linear regression models and effects graphs are presented in Figure 2-2. *Sweetness* was influenced ($p<0.0001$, $r^2 = 0.94$) by the concentrations of sucrose ($\beta = 0.84$, $p<0.0001$), coffee extract ($\beta = -0.27$, $p<0.002$), and milk ($\beta = 0.22$, $p<0.007$), with no significant interaction between variables ($p>0.05$). *Coffee* flavor was dominated ($p<0.0001$, $r^2 = 0.96$) by the concentration of coffee extract ($\beta = 0.95$, $p<0.0001$), but the effect of milk concentration, though smaller ($\beta = -0.16$), was significant ($p<0.01$). Again there were no significant interactions ($p>0.05$). Somewhat surprisingly, *milk* flavor was influenced ($p<0.0001$, $r^2 = 0.87$) most by coffee extract concentration ($\beta = -0.72$, $p<0.0001$), followed by milk concentration ($\beta = 0.52$, $p<0.0002$), with no significant interactions ($p>0.05$). Apparently the strong flavor of coffee masked the more subtle dairy flavor. *Thickness* was significantly influenced ($p=0.0045$, $r^2=0.72$) by milk ($\beta = 0.64$, $p<0.0007$) and sucrose concentration ($\beta = 0.45$, $p<0.0110$), largely through their effect on total solids content (Figure 2-3).

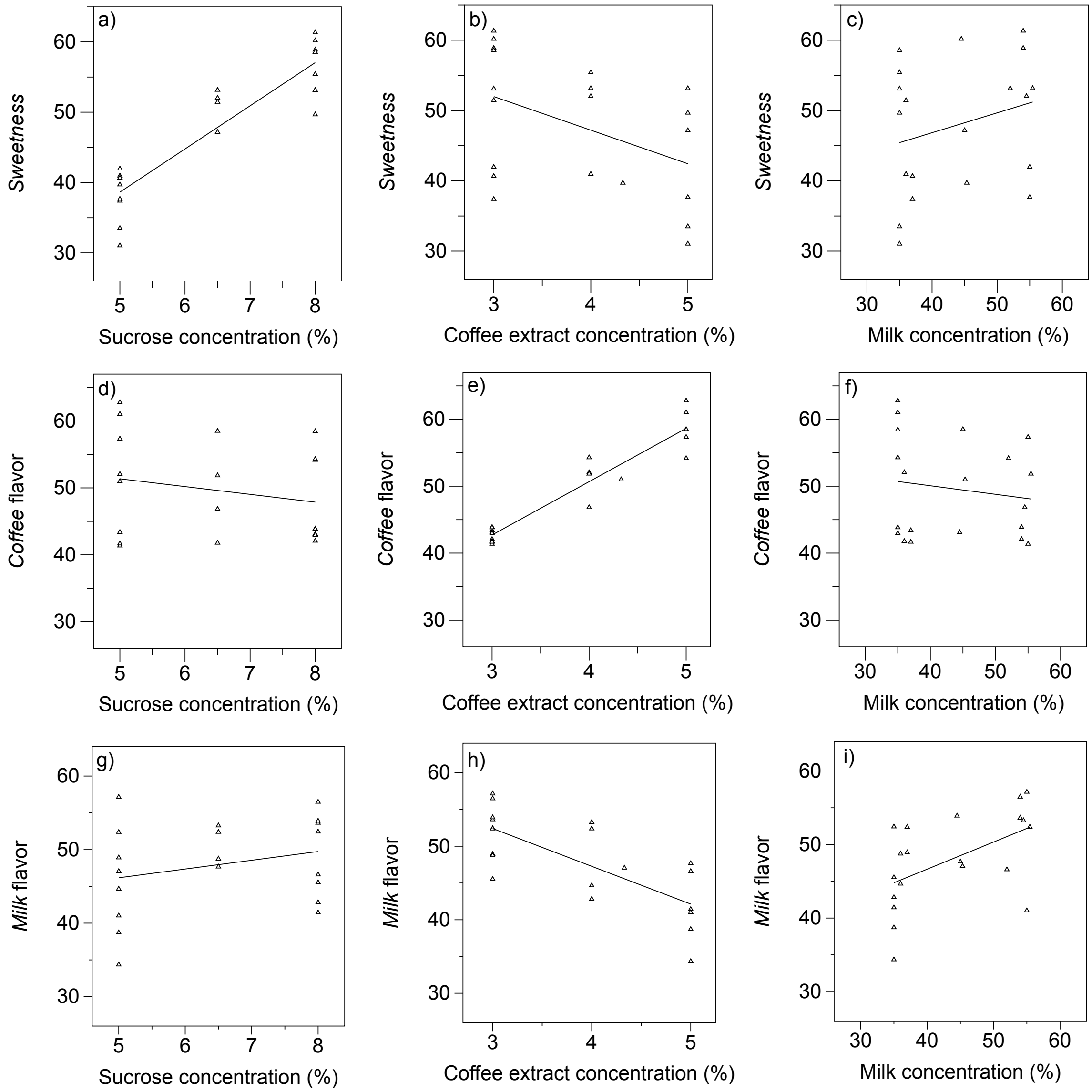

Figure 2-2 Effects graphs for psychophysical models. *Sweetness* (a,b,c), *coffee* (d,e,f) and *milk* (g,h,i) as a function of sucrose (a,d,g), coffee extract (b,e,h) and milk (c,f,i) concentrations.

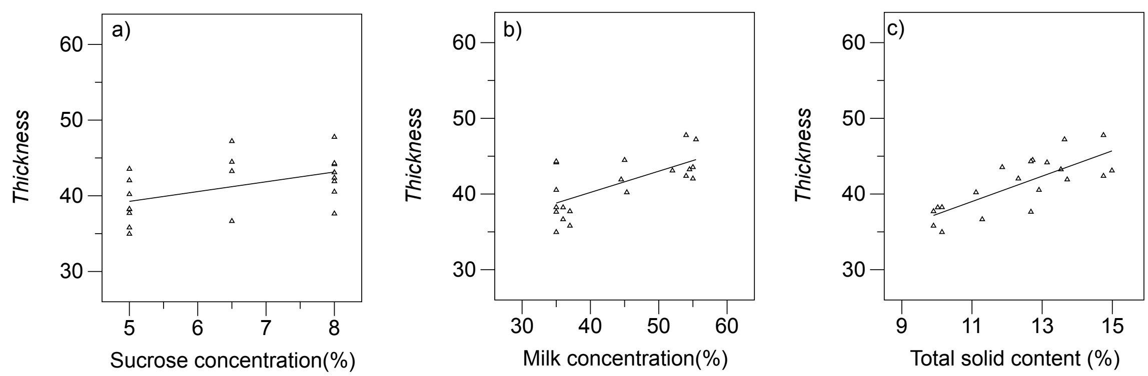

Figure 2-3 Effects graphs for *thickness* as a function of sucrose (a) and milk (b) concentrations and total solids content of the beverages (c).

3.4 Effects of perceived attribute intensities on *liking*

The psychohedonic (sensation-liking) model based on perceived intensity (*sweetness*, *milk* and *coffee*), explained 63% of the variance in *liking* (fitted model: *liking* = 0.03+0.06**milk*+0.03**coffee*+0.03**sweetness,* p=0.001). *Sweetness* (β= 0.53) and *milk* (β= 0.69) contributed significantly to the model (p's <0.04), while *coffee* (β= 0.48) was marginal (p=0.09).

4. Discussion

Previous studies have demonstrated that overall liking scores can be influenced by the inclusion of attribute diagnostic questions on the ballot. Popper, Rosenstock, Schraidt, and Kroll (2004) found that asking participants to rate attribute intensities using JAR scales influenced the average ratings of overall liking, but intensity scales had no such effect. The presence of attribute JAR ratings could increase or decrease the ratings for overall liking (Earthy, MacFie, & Hedderley, 1997). In contrast, attribute intensity rating did not show a significant effect on overall liking (Mela, 1989; Vickers, Christensen, Fahrenholtz, & Gengler, 1993). Based on this, we anticipated that the method used to rate attributes (intensity scales, ideal scaling or JAR scaling) would influence the overall *liking,* but we found no such effect.

In contrast to the physicohedonic (formulation-liking) model where the influence of milk and sucrose were equivalent, *milk* had a larger influence on *liking* than did *sweetness* in the psychohedonic model. More critically, the direction of the effect for *coffee* flavor was opposite that of coffee extract. That is, more *coffee* flavor resulted in greater liking, whereas in the physicohedonic model, more coffee extract either had no effect, or may have even reduced liking slightly. Informal tasting revealed that more coffee extract increased bitterness in addition to *coffee* flavor. However, we failed to ask consumers to rate bitterness as an attribute, and we are thus unable to formally model its effect on liking statistically. However, Boeneke, McGregor, and Aryana (2007) reported that increasing coffee flavor was accompanied by an increase in bitterness of roasted coffee as assessed by a trained panel, and Moskowitz and Gofman (2007) reported consumer liking of coffee increased with the intensity of its bitterness to a maximum, beyond which liking declined. Furthermore, we observed a reduction in *sweetness* with increasing concentrations of coffee extract (Figure 2-2b) as would be expected from suppression of sweetness by bitterness (Lawless, 1979).

In the present case it is likely that *coffee* flavor was perceived as more than just bitterness, or a percept other than bitterness, since *liking* was positively related to *coffee* flavor but negatively related to the amount of coffee extract. This is a subtle, but important, distinction, for if a product developer could increase *coffee* flavor, say through the introduction of key aroma compounds, without a coincident increase in bitterness (as comes with simply adding more coffee extract), then the two percepts could be optimized separately.

The psychohedonic model explained more variance in *liking* than the physicohedonic model, consistent with prior results (Hayes & Duffy, 2008). That perceived intensities are better predictors of *liking* than formulation is entirely expected, as perceived intensity is presumed to mediate the relationship between concentration and *liking*. Milk concentration has been shown to affect the optimal sucrose concentration in coffee (Moskowitz, 1985), and the interaction between milk and coffee flavor (Parat-Wilhelms et al., 2005) was critical to consumer acceptance of a coffee milk beverage (Boeneke et al., 2007). In the present case, our coffee-flavored dairy beverage is a complex food matrix, where perceptual and physical interactions among stimuli are quite common. Adding more sucrose might be expected to reduce the bitterness of coffee (Pangborn, 1982) via mixture suppression that occurs centrally (Lawless, 1979), whereas adding milk fat may alter bitterness via partitioning (see Bennett, Zhou & Hayes, 2012). Previously, Lawless (1977) observed non-intuitive perceptual interactions when attempting to predict liking from stimulus concentration in simple mixtures: adding a small amount of quinine (which is unpleasant by itself) counterinuitively increases liking for a sweet-bitter mixture by reducing excessive sweetness. Here, we confirm similar complex interactions and extend them beyond model systems to a real food product.

5. Conclusions

A psychohedonic (intensity-liking) model is better than physicohedonic (formulation-liking) model for predicting consumer liking. However, a physicohedonic model might be more actionable from a product formulation perspective. Product developers and sensory specialists should always remember that a single ingredient may influence more than one perceptual attribute, especially in a complex food. Unlike some pure compounds, e.g. sucrose, which might uniquely produce a single *sweetness* perception, adding more coffee extract not only increases coffee flavor but also bitterness. Psychophysical models can help in understanding and interpreting the results from physicohedonic and psychohedonic models.

6. Acknowledgments

This project was partially supported by NIH Grant AI094514 to JEH and GRZ. The authors thank Hanna Schuster for preliminary formulations and Maggie Harding for preparing the coffee milk samples. We also thank the Sensory Evaluation Center staff for their assistance with this test.

Chapter 3

Just-About-Right and ideal scaling provide similar insights into the influence of sensory attributes on liking

Submitted to Food Quality and Preference

Abstract

Just-About-Right (JAR) scaling has been criticized for measuring attribute intensity and acceptability simultaneously. Using JAR scaling, an attribute is evaluated for its appropriateness relative to one's hypothetical ideal level that is pre-defined at the middle of a continuum. Alternatively, ideal scaling measures intensity and acceptability separately. Ideal scaling allows participants to rate their ideal freely on the scale (i.e., without assuming the "Too Little" and "Too Much" regions are equal in size). We hypothesized that constraining participants' ideal to the center point, as is done in the JAR scale, may cause a scaling bias and, thereby, influence the magnitude of "Too Little" and "Too Much" estimates. Furthermore, we hypothesized that the magnitude of "Too Little" and "Too Much" would influence liking to different extents.

Coffee-flavored dairy beverages (n=20) were formulated using a mixture design that varied the ratios of water, milk, coffee extract, and sucrose. Participants tasted 4 of 20 prototypes that were served in a monadic sequential order using an incomplete block design. Participants were randomly assigned to one of three research conditions, two of which are discussed here: ideal scaling (n=129) or JAR scaling (n=132). For both conditions, participants rated overall

liking using a 9-point Quartermaster hedonic scale. Four attributes (*sweetness, milk flavor, coffee flavor* and *thickness*) were evaluated. The reliability of an individual participant's ideal rating for an attribute was assessed by standard deviation of ideal ratings. All data from a participant were eliminated from further analyses if his/her standard deviation of the ideal ratings for any attribute was identified as a statistical outlier. Multiple linear regression was used to model *liking* as a function of "Too Little" or "Too Much" attribute intensities.

All mean ideal ratings were significantly different from the central point of the scale (i.e., 50). *Coffee flavor* (57.2) was the only attribute for which mean ideal rating fell outside the central 10% (45.0-55.0). Contrary to our hypothesis, the magnitude of "Too Little" and "Too Much" was not affected by scaling method. As expected, the influence of the magnitude of "Too Little" and "Too Much" on *liking* was asymmetrical. Both scaling methods agreed that *sweetness* and *coffee flavor* were the main sensory attributes affecting *liking*. Overall, JAR scaling and ideal scaling were comparable for measuring "Too Little" and "Too Much", and identifying the main factors that affected liking.

1. Introduction

Just-About-Right (JAR) scaling is widely applied in the food industry for product development (Popper & Gibes, 2004; Rothman & Parker, 2009; Xiong & Meullenet, 2006; Hayes, Raines, DePasquale, & Cutter, 2014). JAR scales are popular in marketing and R&D departments in the industry due to their ease of use and directional guidance (Ares, Barreiro, & Giménez, 2009; Gacula, Rutenbeck, Pollack, Resurreccion, & Moskowitz, 2007; Popper & Kroll, 2005). JAR scales are a rapid way to determine if an attribute's intensity is at an optimal level (Lawless & Heymann, 2010; Moskowitz, 2001; Popper & Kroll, 2005). This technique is commonly used at an early stage of product development (Pangborn, Guinard, & Meiselman, 1989), when a systematic solution (e.g., full formulation design) is not available, or the cost or time is a matter of concern.

The JAR scale is a bipolar measurement. In JAR scaling, two semantically opposite anchors, e.g., "Not Sweet At All" and "Much Too Sweet", are placed at each end of the scale, and the midpoint is labeled "Just About Right" or "Just Right" (Booth, Thompson, & Shahedian, 1983; Rothman & Parker, 2009; Shepherd et al., 1989; Vickers, 1988). "Just About Right" or "Just Right" is assumed to be a participant's ideal level (van Trijp et al., 2007). Using JAR scaling, an attribute is evaluated for its performance (appropriateness) relative to this ideal level (Rothman & Parker, 2009; Worch et al., 2010). Generally, "Too Little" or "Too Much" attribute intensity is estimated by the deviation of the rating from the center point of the scale. The intensity of an attribute can be increased if it is perceived as "Too Little". On the other hand, the intensity can be decreased if it is perceived as “Too Much”. For this reason, JAR scaling is recognized as a directional tool (Moskowitz, 2001). However, the magnitudes of “Too Little” or “Too Much” does not indicate the size of the intensity change is needed.

JAR scaling combines the measurements of attribute intensity and consumer acceptability (Moskowitz et al., 2008). Some researchers have criticized this practice, and suggested JAR scaling should not replace traditional experimental design for product optimization (Stone & Sidel, 2004). Others claim that JAR scaling is a challenging task for naïve consumers because these ratings involve at least three decisions: a) perception of the attribute intensity; b) location of the participants’ ideal point; and c) comparison of the difference between perceived intensity and ideal point (Moskowitz, 2001; van Trijp et al., 2007). Further, studies find optimal formulations achieved by JAR scaling differ from those predicted by hedonic scores (Epler et al., 1998; Shepherd et al., 1989; Vickers, 1988).

Additionally, JAR scales may incorporate some unique biases. JAR ratings may be influenced by cognitive factors (Rothman & Parker, 2009). For example, a participant who is on a diet may treat “sweetness” of ice cream as a negative attribute, and tend to always rate ice cream as “too sweet”. Conversely, for a product attribute that positively influences *liking*, a participant might always rate it “not enough”. For instance, a participant who likes meat may always rate the meat topping on a pizza “not enough”.

Alternatively, ideal scaling separates the measurements of attribute intensity and acceptability into two separate scales (Gilbert et al., 1996; Rothman

& Parker, 2009; van Trijp et al., 2007; Worch et al., 2012a). In ideal scaling, acceptability is presumably maximized at the ideal intensity level. Unlike JAR scaling, where the ideal level is fixed at the central point of the scale, ideal scaling allows a participant to designate his/her hypothetical ideal level anywhere on the scale, and “Too Little” and “Too Much” are estimated by the difference between perceived intensity and ideal intensity. Ideal scaling has been applied in the industry and academia for decades (Gilbert et al., 1996; Goldman, 2005; Hoggan, 1975; van Trijp et al., 2007; Worch et al., 2012a). However, comparisons of JAR scaling and ideal scaling for measurement of “Too Little” or “Too Much” are lacking. Here we hypothesized that participants’ ideal intensities would differ from the central point of the scale, which consequently may influence the measurement of “Too Little” and “Too Much”, and their influence on liking.

2. Materials and methods

This study was a part of a larger experiment designed to optimize a coffee-flavored dairy beverage for a retail facility on the Penn State campus. Participants (n=388 in total) were randomly assigned to one of three research conditions that differed only in ballot design. For the purpose of this analysis, only the data from research conditions that applied ideal scaling and JAR scaling are discussed here. In both conditions, participants were asked to rate *liking* as well as attribute intensities for *sweetness*, *milk flavor*, *coffee flavor* and *thickness*. Procedures were exempted from IRB review by the Penn State Office of Research Protections staff under the wholesome foods exemption in 45 CFR

46.101(b)(6). Participants provided implied informed consent and were compensated for their time.

2.1 Subjects

A total of 261 participants (70 men) completed the product evaluation described here using either ideal scaling (n=129) or JAR scaling (n=132). Participants were recruited ahead of time using an existing participant database managed by the Sensory Evaluation Center at Penn State, or via staff intercepts in public spaces in or around the Food Science Department at Penn State.

To qualify for participation, individuals had to be regular drinkers of coffee or coffee-flavored beverages (Table 3-1), and free of food allergies. The majority of participants (105) were between 18-27 years old, 49 were 28-37, 38 were 38-47, 48 were 48-57, 18 were 58-67, and only 3 were over 67 years old. The majority were white (n=205, ~78.5%), 36 identified themselves as Asian or Pacific Islander, 7 as African or African American, 8 as Hispanic/Latino, and 5 did not report their ethnicities.

Table 3-1. Frequency (%) of regularly consumed coffee-flavored beverages

Products	Ideal scaling (n=129)	JAR scaling (n=132)
Cappuccino	20.9	30.3
Latte	24.0	37.9
Black Coffee	27.9	25.8
Iced Coffee	25.6	37.9
Coffee with milk, cream, and/or sucrose	61.2	58.3

Note: This is a "check all that apply" question, so the sums in a column may exceed 100%.

2.2 Sample formulation and preparation

Coffee-flavored dairy beverages (n=20) were formulated using a mixture design with four constrained variables: coffee extract (3.0-5.0 wt %; Autocrat Sumatra 1397, Autocrat Natural Ingredients, Lincoln, RI), sucrose (5.0-8.0 wt %), milk (35-55 wt %, 2% fat), and water (35-55 wt %). These components accounted for 99.8% of the individual formulations. A constant amount of pectin (0.2 wt %; Grinsted® SY, Dupont Danisco) was added to all the samples. The exact composition of each formula is shown in Table 3-2.

Table 3-2. Sample formulations (in weight percentage)

Product[1]	Milk (%)	Water (%)	Coffee extract (%)	Sucrose (%)	Solid content (%)[2]
1	35.93	54.89	3.99	4.99	10.02
2	45.24	45.24	4.32	4.99	11.11
3, 19	36.93	54.89	2.99	4.99	9.90
4, 9	53.89	34.93	2.99	7.98	14.75
5	34.93	51.90	4.99	7.98	13.14
6, 18	34.93	54.89	4.99	4.99	10.15
7	44.91	43.41	4.99	6.49	12.73
8	35.93	54.39	2.99	6.49	11.29
10	54.89	34.93	4.99	4.99	12.32
11	44.41	44.41	2.99	7.98	13.71
12, 16	54.39	34.93	3.99	6.49	13.53
13	54.89	36.93	2.99	4.99	11.86
14, 17	34.93	53.89	2.99	7.98	12.68
15	51.90	34.93	4.99	7.98	14.99
20	34.93	52.89	3.99	7.98	12.91

[1]Samples in the same row share the same formulation.

[2]Calculated from the solids content of the ingredients.

For convenience, pectin was mixed with sucrose completely before blending them with water, milk, and coffee extract to make sample batches (Figure 1-1). Batches were heated up to 72º C and held 15 seconds to assure that the sucrose was completely dissolved, the pectin dispersed and the product was safe for consumption. The finished samples were kept at refrigeration temperature (~4.5°C) for at least 24 hours before serving. Two ounces of coffee milk were served in 4-oz Solo transparent plastic cups (Solo Cup Company, Urbana, IL).

2.3 Sensory evaluation

Participants were randomly assigned to a research condition upon entering the test booths. To minimize fatigue, an incomplete block design (Gacula, 2008a) was applied, i.e., each participant tasted only 4 of 20 samples. For each sample, participants were asked to rate their overall *liking* and attribute intensity. The attributes assessed included *sweetness*, *milk flavor, coffee flavor*, and *thickness*.

Liking was assessed using a standard 9-point Quartermaster hedonic scale (1="Dislike Extremely", 5 ="Neither Like Nor Dislike", and 9="Like Extremely") (Peryam & Pilgrim, 1957). Attribute intensities, both perceived and ideal, were measured using continuous line scales (0-100); descriptive anchors were placed at 10% and 90% of these scales, representing low intensity (e.g., "Not At All Sweet") and high intensity (e.g., "Extremely Sweet"). Just-About-Right (JAR) scales were designed as continuous line scales with three descriptive anchors, low intensity (i.e., “Much Too Weak”) on the left end, “Just About Right” at the center, and high intensity (i.e., “Much Too Strong”) on the right end. Demographics and consumption behavior for coffee-based beverages were collected after all samples were evaluated.

The ballot was administered and data were collected using Compusense *five*® software (Compusense Inc., Ontario, Canada). Samples were served in a serial monadic order, with a two-minute mandatory break between each sample. Participants were asked to rinse with room temperature filtered water between samples to reduce potential carry-over effects.

2.4 Data analysis

Data analyses were carried out using the statistical software JMP® (v9.02, SAS Institute Inc.). Significance criteria were set to α=0.05.

“Too Little” and “Too Much” refer to perceived intensity rated below or above either the ideal intensity in the ideal scale or the “Just About Right” point in the JAR scale. “Too Little” and “Too Much” were calculated as the distance between the actual rating and ideal level (i.e., ideal intensity or “Just About Right” point). The reliability of ideal ratings for an individual participant was evaluated using the standard deviations (n=4) of ideal ratings for an attribute. Outliers were identified using Tukey’s box-and-whisker plot. All data from an individual participant were eliminated from further analyses when the standard deviation of any attributes’ ideal ratings (*sweetness, milk flavor, thickness, and coffee flavor*) for that individual was identified as an outlier.

After outliers were removed from the data, the stability of ideal ratings was assessed by the effect of product using analysis of variance (ANOVA), where participant was a random effect and both product and serving order were treated as fixed effects. The average of self-reported ideal intensities for *sweetness*, *milk flavor, coffee flavor*, and *thickness* was compared to the central point (i.e., 50) of a line scale using a t-test.

For ideal scaling, the mean of ideal intensities (n=4) of an attribute for an individual consumer was calculated and applied as the ideal intensity level for the calculation of “Too Little” and “Too Much” for that attribute for that individual participant. To investigate the effect of scaling method on the magnitudes of “Too Little” and “Too Much”, analysis of variance (ANOVA) was applied. In this ANOVA model, the participant was considered a random effect nested within the scaling method (scale), and product and scale were considered as fixed effects; the interaction of product by scale was also included in the model. For convenience of interpretation, “Too Little” was negative and “Too Much” positive in this analysis. For both scaling methods, multiple linear regressions were used to evaluate the effect of “Too Little” and “Too Much” on *liking* (Li, 2011; Worch et al., 2010). In the regression models the absolute values of “Too Little” and “Too Much” were used.

3. Results

3.1 Reliability (individual) and stability (panel) of ideal ratings

The reliability of individual participants' ideal ratings was assessed using Tukey's box-and-whisker plots of standard deviations of their ratings (Figure 3-1). Except for one participant (ID=50) who precisely indicated his/her ideal level for each attribute (i.e., standard deviations of 0), participants showed variance in their ideal ratings for all the attributes. Several individuals were identified as outliers. As a result, data from 15 out of 129 participants were excluded from further analyses.

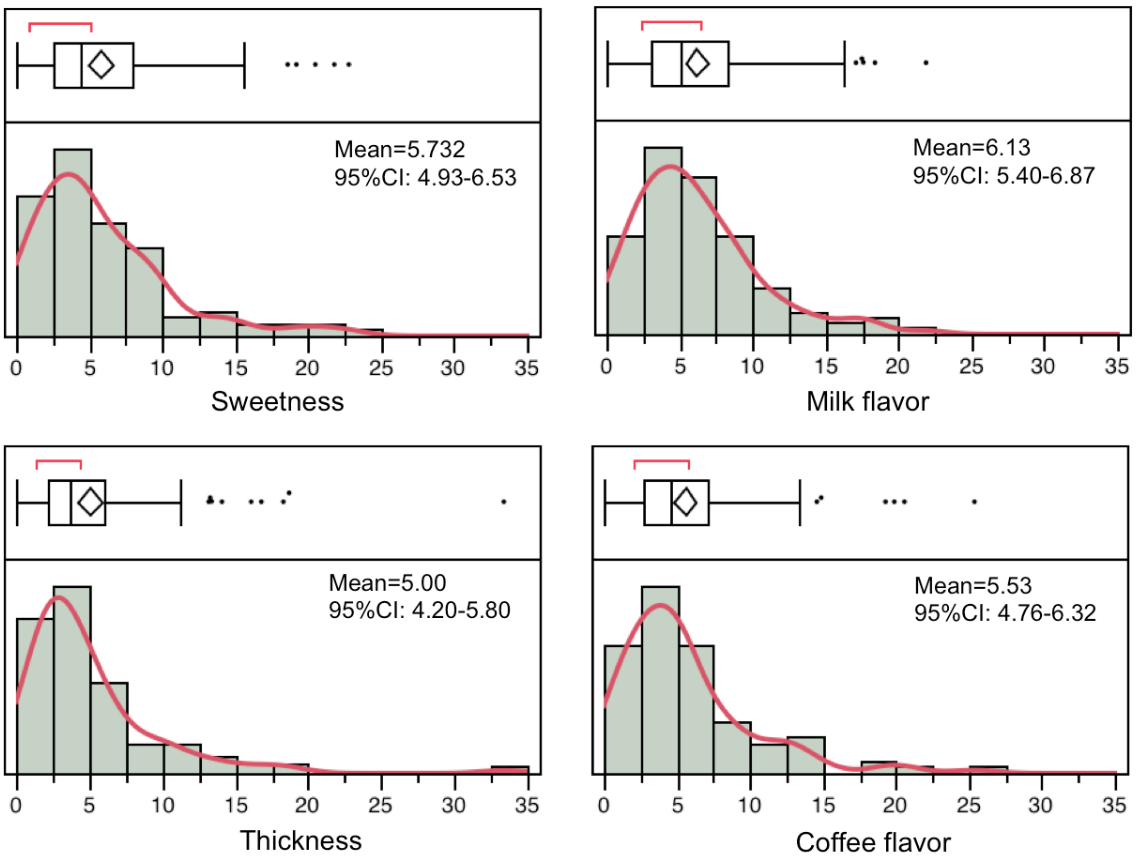

Figure 3-1. Distribution of standard deviations of individual ideal ratings

Data in the Tukey's box-and-whisker plots beyond the terminus of the whisker were identified as outliers.

The stability of ideal intensity ratings was investigated by evaluating the effect of product using ANOVA, similar to Worch and Ennis (2013). All ANOVA models had adjusted R-squares that were greater than 85% (Table 3-3). Product showed a marginal effect on ideal ratings of *sweetness*. Product did not show a significant effect on ideal ratings of other attributes. Notably, serving order, i.e., 1st, 2nd, 3rd, or 4th, significantly influenced ideal ratings for *coffee flavor*, although, means of ideal ratings for *coffee flavor* by serving order varied by less than 2% of the scale (1st=58.7, 2nd=57.1, 3rd=56.33, and 4th=56.4).

Table 3-3. Effect of product and serving order on ideal ratings

Attribute	Effect	F-Ratio	p-value	R^2-adj.
Sweetness	**Product**	**1.621** $\mathbf{_{(19,\ 329.5)}}$	**0.0493**	87.30%
	Serving order	1.471 $_{(3,\ 319.9)}$	0.2225	
Milk flavor	Product	1.102 $_{(19,\ 330.1)}$	0.3469	86.90%
	Serving order	0.564 $_{(3,\ 320.3)}$	0.6391	
Thickness	Product	1.372 $_{(19,\ 327.4)}$	0.1381	90.10%
	Serving order	0.921 $_{(3,\ 320.3)}$	0.4308	
Coffee flavor	Product	0.699 $_{(19,\ 328.1)}$	0.8192	89.10%
	Serving order	**4.582** $\mathbf{_{(3,\ 320.0)}}$	**0.0037**	

Note: p-values in bold indicate terms are significant in the model at α=0.05.

3.2 Distribution characteristics of ideal intensity ratings

In the ideal scaling condition, participants were allowed to rate their ideal intensities anywhere on the scale. The distributions of mean ideal intensities for the attributes are illustrated in Figure 3-2. Across the group, participants used almost the full range of the line scale to rate their ideal intensities. The means of ideal intensities (for each participant n=4) for all four attributes had wide variations with standard deviations greater than 10. Except for *milk flavor* that has an overall mean of 47.4 ($p=0.0567$), overall means of ideal intensities for the other attributes were significantly different from the central point of the scale (i.e., 50). Ideal *sweetness* had a mean rating of 47.4 ($p=0.0477$). *Thickness* had a mean ideal of 45.2 ($p<0.0001$), and the mean ideal *coffee flavor* had a mean ideal of 57.2 ($p<0.0001$).

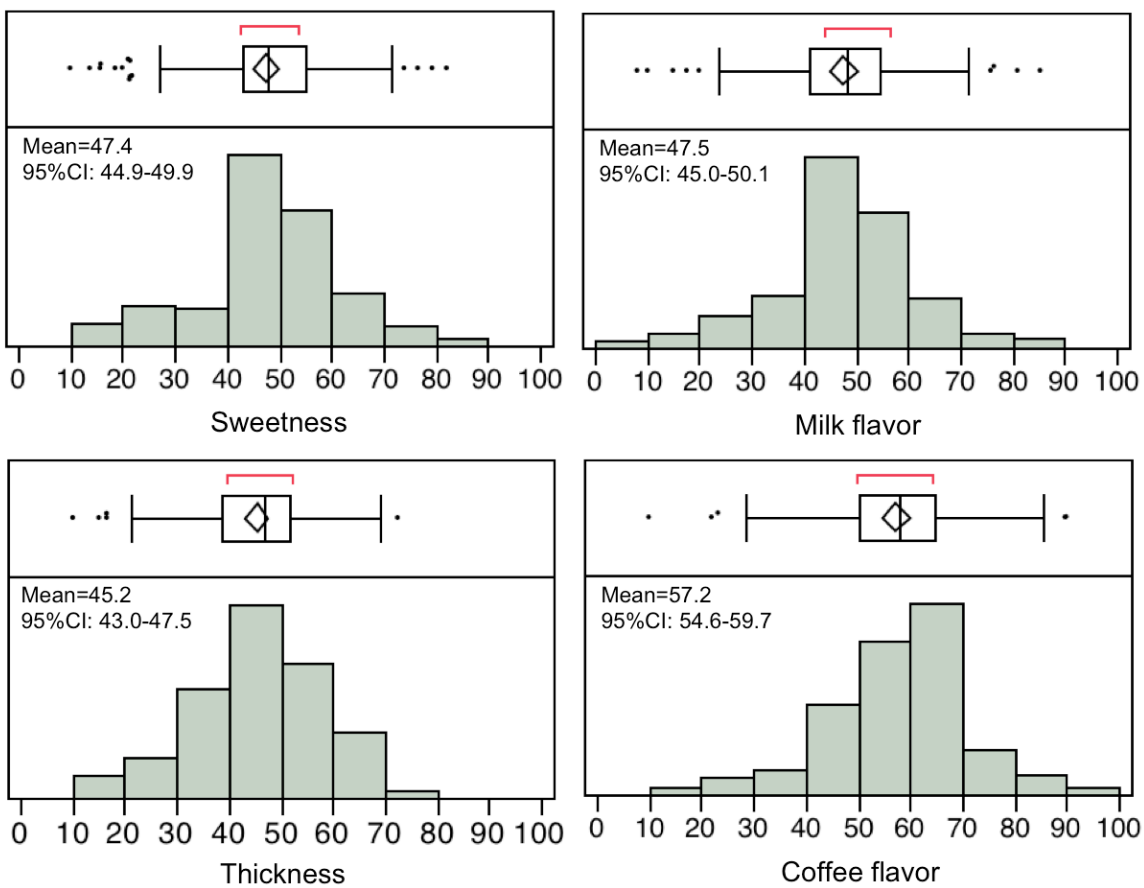

Figure 3-2. Distributions of ideal intensity ratings using the mean ideal for each participant (n=114)

3.3 Influence of scaling method on "Too Little" and "Too Much"

Generally, means of ideal intensities were different from the central points of the ideal scales, i.e., 50. However, the magnitude of these differences was <10% of the full scale range. Since "Too Little" and "Too Much" were defined as the deviations of perceived intensities from ideal intensities, the resulting asymmetry of ideal rating may influence "Too Little" and "Too Much" estimates. Therefore, "Too Little" and "Too Much" were compared between the two scaling methods (Table 3-4 and Figure 3-3).

None of the interaction terms (product by method) were significant. As expected, product had a significant effect on "Too Little" and "Too Much" for most attribute intensities, with the exception of "Too Much" *thickness,* which is reasonable given that all these prototypes were formulated with the same amount of pectin. In contrast, scaling method did not have a significant effect on "Too Little" and "Too Much" for any attribute ($p>0.05$).

Table 3-4. Effects of product and method on "Too Little" and "Too Much"

Performance	Attribute	Term	F-Ratio	p-value	R^{2}-adj.
Too little	*Sweetness*	**Product**	$\mathbf{9.53_{(19,517.5)}}$	**<.0001**	62.4%
		Method	$0.11_{(1,229.9)}$	0.741	
		Product*Method	$0.73_{(19,517.5)}$	0.7938	
	Milk flavor	**Product**	$\mathbf{2.17_{(19,323.0)}}$	**0.0035**	53.9%
		Method	$0.72_{(1,169.9)}$	0.3974	
		Product*Method	$1.42_{(19,323.0)}$	0.1139	

	Coffee flavor	**Product**	$\mathbf{2.63_{(19,461.8)}}$	**0.0002**	53.3%
		Method	$3.57_{(1,221.0)}$	0.06	
		Product*Method	$1.09_{(19, 461.8)}$	0.3605	
	Thickness	**Product**	$\mathbf{2.86_{(19,456.8)}}$	**<.0001**	50.4%
		Method	$0.48_{(1,193.9)}$	0.4884	
		Product*Method	$1.20_{(19,456.8)}$	0.251	
Too much	*Sweetness*	**Product**	$\mathbf{3.72_{(19,136.4)}}$	**<.0001**	79.1%
		Method	$1.36_{(1,149.8)}$	0.2461	
		Product*Method	$1.02_{(19,136.4)}$	0.4446	
	Milk flavor	**Product**	$\mathbf{2.74_{(19,340.2)}}$	**0.0002**	42.6%
		Method	$0.79_{(1,160.7)}$	0.3762	
		Product*Method	$1.03_{(19,340.2)}$	0.4281	
	Coffee flavor	**Product**	$\mathbf{3.34_{(19,217.7)}}$	**<.0001**	64.4%
		Method	$1.23_{(1,179.5)}$	0.2695	
		Product*Method	$1.05_{(19,217.7)}$	0.402	
	Thickness	Product	$0.99_{(19,185.1)}$	0.478	54.1%
		Method	$0.35_{(1,140.3)}$	0.5528	
		Product*Method	$1.48_{(19,185.1)}$	0.0981	

Note: p-values in bold indicate terms are significant in the model at α=0.05.

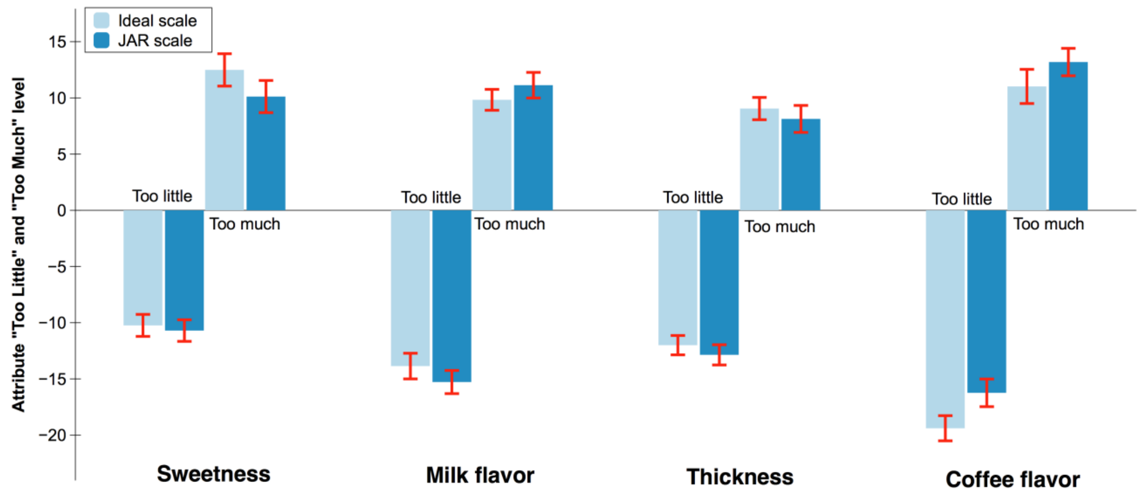

Figure 3-3. Attribute “Too Little” and “Too Much” comparison between two scaling methods

3.4 Influence of “Too Little” and “Too Much” on *liking*

Prior work suggests that “Too Little” and “Too Much” of a sensory attribute might impact overall liking differently (Vickers, Holton, & Wang, 1998; Xiong & Meullenet, 2006). In other words, consumers show different tolerances for “Too Little” and “Too Much”. In this section, the effects of “Too Little” and “Too Much” on *liking* were investigated through multiple linear regression by fitting *liking* as a function of “Too Little” and “Too Much” (Figure 3-4).

For ideal scaling, 32.9% of variation in *liking* was explained by the multiple linear regression model ($F_{8, 447}$=26.14, p<0.0001). Except for “Too Much” *milk flavor* (p=0.2555), and “Too Little” (p=0.5266) and “Too Much” (p=0.0906) *thickness,* which were not significant, all other attributes (“Too Little” or “Too Much”) showed significant influence on *liking*. For ideal scaling, “Too Little” *sweetness* had the strongest impact on *liking*, followed by “Too Much” and “Too Little” *coffee flavor*.

For JAR scaling, the regression model explained 45.9% of variation in *liking* ($F_{8, 519}$=56.87, p<0.0001). “Too Little” and “Too Much” for all attributes significantly affected *liking*. Consistent with the results for the ideal scaling above, “Too Little” *sweetness* had the strongest impact on *liking,* followed by “Too Much” and “Too Little” *coffee flavor*. The impact of *milk flavor* on *liking* seemed more symmetrical for the JAR scale.

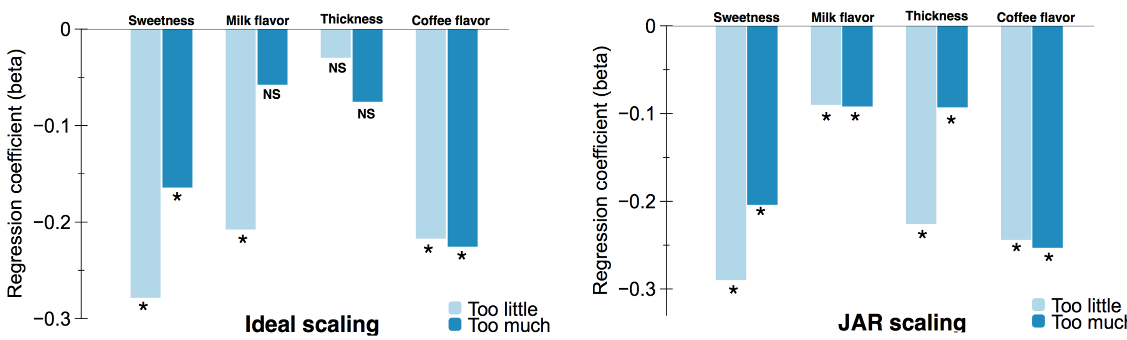

Figure 3-4. Influence of attribute “Too Little” and “Too Much” on *liking*

Notes: 1. Bars marked with * indicate attribute showed significant influence on liking at α=0.05. 2. NS showed a non-significant influence on liking.

4. Discussion

4.1 Reliability and stability of ideal ratings

Both JAR scaling and ideal scaling measure attribute performance using the concept of a participant’s “ideal”. However, some researchers question whether participants can have an abstract concept of their ideal except in relation to a physical sample (Moskowitz et al., 2008; Rothman & Parker, 2009).

Participants are assumed to have an implicit ideal point in their mind (Popper & Gibes, 2004), and are expected to rate their ideal precisely on the ideal scale (Worch, Le, Punter, & Pages, 2013). Several studies have shown that participants are highly reliable in rating their ideal intensities (Goldman, 2005; Mcbride & Booth, 1986; van Trijp et al., 2007; Worch, Le, Pages, 2010) for those attributes that are well understood. However, ideal ratings might show some variance when participants do not understand attributes well. To avoid potential misinterpretation, checking the reliability of ideal ratings is strongly recommended (Worch et al., 2012a).

Standard deviation is useful for evaluating the reliability of panelist's ratings (Mandel, 1991; Meilgaard, Civille, & Carr, 2007; Rossi, 2001). Here, after the exclusion of statistical outliers, the standard deviations for all ideal intensities were less than 16.0 (16% of the scale range), and 90% of these standard deviations were lower than 10.0. To our knowledge, there are no published guidelines for evaluating the stability of ideal ratings using standard deviations and scaling range in the literature. However, it has been reported that even a well-trained descriptive panel will have standard deviations around 10% of scale range for attribute intensity ratings (Lawless, 1988). Consumer panels generally perform ever worse in attribute intensity ratings; variation may reach more than 25% of the scale range (Lawless & Heymann, 2010). We conclude that our participants (naïve consumers) overall showed good reliability in their ideal ratings.

The stability of ideal ratings for the whole panel was evaluated through the effect of product (Worch, Le, Punter, & Pages, 2012b). Product showed a marginally significant effect only for ideal *sweetness* (p=0.0493). This effect was due to one sample (sample #6, Table 3-2). Since its replicate (sample #18, Table 3-2) was not significantly different from the others, we suspect this result may reflect Type I error. Means of ideal *sweetness* were not significantly different across products when this sample (sample #6) was eliminated from the dataset. Ideal ratings across serving orders were also investigated and compared. *Coffee flavor* was the only attribute whose ideal ratings significantly differed across serving order. Interestingly, the mean values of ideal *coffee flavor* seemed to decrease with order (1^{st}=58.7, 2^{nd}=57.1, 3^{rd} =56.3, and 4^{th}=56.4), which might indicate sensory fatigue and adaption during evaluation or bitterness built across evaluations. Nonetheless, this slight difference may not be of practical importance. In general, the ideal ratings of panel performance were stable.

4.2 “Too Little” and “Too Much” between JAR scaling and ideal scaling

JAR scaling and ideal scaling differ in how they define the ideal level on the scale. In ideal scaling, participants used nearly the entire range of the scale for their ideal ratings. In contrast, constraining a subject’s ideal level at the central point of the scale as is done in JAR scaling may be expected to introduce bias. However, contrary to our hypothesis, we observed no effect of scaling method on the magnitude of “Too Little” and “Too Much” estimates. JAR scaling and ideal scaling appear to be highly comparable in measuring attribute “Too Little” and “Too Much” intensities, at least for present data.

Even though this similarity was observed between the two scaling methods, some differences are notable. All the “Too Little” and “Too Much” scores significantly influenced *liking* in JAR scaling. In ideal scaling, “Too Much” *milk flavor*, and both “Too Little” and “Too Much” *thickness* did not show significant impact on *liking* (Figure 3-4). In addition, in the multiple linear regression models, the JAR scaling model (45.9%) explained more variance in *liking* than the ideal scaling model (32.8%). These findings indicate attribute “Too Little” and “Too Much” estimated by the JAR scaling may better predict liking when compared to values obtained from ideal scaling. Currently it is unknown which scaling would be more valid for detecting attribute impacts on consumer liking. Therefore, further studies are warranted.

4.3 Asymmetrical influence of attribute “Too Little” and “Too Much” on *liking*

With both the JAR scaling and ideal scaling, the attribute “Too Little” and “Too Much” affected *liking* asymmetrically (Figure 3-4). Participants showed different tolerance levels for deviation from their ideals depending on whether they were “Too Little” or “Too Much”. This result agrees with prior reports (Moskowitz, 2001; Xiong & Meullenet, 2006). Both scaling methods agree *sweetness* and *coffee flavor* were more important factors for consumer *liking* as compared to *milk flavor* and *thickness*. This finding matches our expectations about the importance of *sweetness* and *coffee flavor* for *liking* over a coffee-flavored dairy beverage. Both scaling methods agreed that “Too Little” *sweetness* had the highest negative impact on *liking*, followed by “Too Much” and “Too Little” *coffee flavor*, and “Too Much” *sweetness*. The asymmetric impacts of “Too Little” and “Too Much” on *liking* varied across attributes. The asymmetry is greater for *sweetness* than that for *coffee flavor* (Figure 3-5). The classic inverted U shaped relationship between liking and attribute intensity reported in the literature (e.g. Keast & Hayes, 2011; Moskowitz, 1971; Pfaffmann, 1980) may really be an “L”. Here, *sweetness* of a coffee-flavored dairy beverage seemed to fit this pattern well.

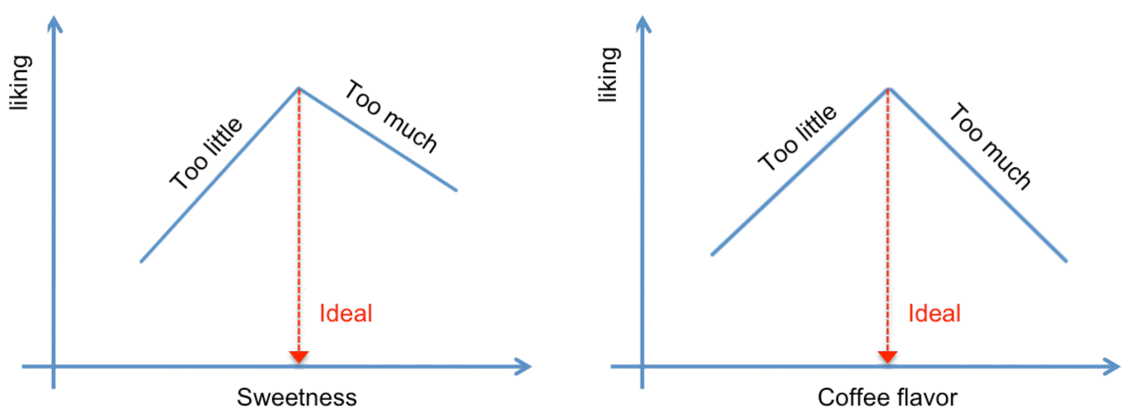

Figure 3-5. Asymmetrical impacts of *sweetness* and *coffee flavor* on *liking*

Compared to “Too Much”, “Too Little” *sweetness* showed a higher impact on *liking*. This means consumers preferred a coffee-flavored dairy beverage to be “Too Sweet” rather than “Not Sweet Enough” when an ideal *sweetness* was not achievable. This is similar to a yogurt study, where “Too Much Sweet” was less harmful to *liking* than “Not Sweet Enough” (Vickers, Holton, & Wang, 2001).

This finding is very meaningful for product development, as it is less risky to make a coffee-flavored dairy beverage “Too Sweet” rather than “Not Sweet Enough”. On the surface, the effect of *coffee flavor* on *liking* is contradictory to our understanding that *coffee flavor* is a positive factor for consumer liking as “Too Much” *coffee flavor* had a slightly higher impact on *liking* than “Too Little” *coffee flavor*. However, in addition to increasing *coffee flavor,* adding more coffee syrup also increased bitterness, though we did not ask panelists to rate this attribute. Bitterness is generally regarded as a negative factor to consumer *liking.* For additional discussion, see our previous comparison of psychohedonic and physicohedonic models (Li, Hayes, & Ziegler, 2014).

5. Conclusions

Sweetness and *coffee flavor* were two critical sensory attributes that influence consumer acceptability for a coffee-flavored dairy beverage. “Too Much” *sweetness* had less negative affect on consumer liking than “Too Little” *sweetness*. Thus, it is less risky for a product developer to have a “too sweet” coffee-flavored dairy beverage than one that is “not sweet enough”; whether this generalizes to other products is unknown, but the yogurt data from Vickers’ group suggests it might (Vickers, Holton, & Wang, 2001). Coffee extract is a complex ingredient. Adding more coffee extract into a coffee-flavored dairy beverage might also inevitably produce some negative attribute, like bitterness, which would negatively impact *liking* (Li et al. 2014). Therefore, the level of coffee extract for a coffee-flavored dairy beverage should be selected carefully to balance positive and negative sensory attributes.

Even though JAR scaling and ideal scaling differ in how they place a participant's ideal level on the scale, both scales provided similar estimates of "Too Little" and "Too Much" attribute intensities. Both scaling methods were equally efficient in identifying the main sensory factors that affected consumer liking for a coffee-flavored dairy beverage. This result further justifies the use of JAR scaling for product optimization, which also was found in other studies (Lovely & Meullenet, 2009; van Trijp et al., 2007). By avoiding noise in the rating of attribute ideals and the greater time required with ideal scaling (dual ratings for each attribute), JAR scaling is recommended for product optimization due to increased efficiency.

6. Acknowledgments

This project was partially supported by NIH Grant AI094514 to JEH and GRZ. The authors thank Hanna Schuster for preliminary formulations and Maggie Harding for preparing the coffee milk samples, and thank the Sensory Evaluation Center staff for their assistance with this test. We also thank Dr. Emma Feeney for her critical comments on this manuscript.

Chapter 4

Product optimization: minimizing attributes “Too Little” or “Too Much” is not equivalent to maximizing overall liking

Submitted to Food Quality and Preference

Abstract

In Just-About-Right (JAR) scaling and ideal scaling, attribute delta (i.e., “Too Little” or “Too Much”) reflects a subject’s dissatisfaction level for an attribute relative to their hypothetical ideal. Dissatisfaction (attribute delta) is a different construct from consumer acceptability measured via liking. Therefore, we hypothesized minimizing dissatisfaction and maximizing liking would yield different optimal formulations. The objective of this paper was to compare product optimization between maximizing liking and with minimizing dissatisfaction.

Coffee-flavored dairy beverages (n=20) were formulated using a fractional mixture design that constrained the proportions of coffee extract, milk, sucrose, and water. Participants (n=388) were randomly assigned to one of three research conditions, where they evaluated 4 of the 20 samples using an incomplete block design. Samples were rated for overall *liking* and for intensity of the attributes *sweetness*, *milk flavor*, *thickness* and *coffee flavor*. When appropriate, *Ideal_Delta* and *JAR_Delta* were calculated as the sum of the four attribute deltas as a measure of overall product quality. Optimal formulations were estimated by: a) maximizing *liking*; b) minimizing *Ideal_Delta* or; c) minimizing

JAR_Delta. A validation study was conducted to evaluate product optimization models.

Participants stated a preference for a coffee-flavored dairy beverage with more coffee extract and less milk and sucrose in the dissatisfaction model when compared to the formula obtained by maximizing *liking*. That is, when *liking* was optimized, participants generally liked a “weaker, milkier and sweeter” coffee-flavored dairy beverage. These predictions were verified in a validation study. These findings are consistent with the view that JAR and ideal scaling methods both suffer from attitudinal biases that are not present when liking is rated (i.e., consumers sincerely believe they want ‘dark, rich, hearty’ coffee when they do not).

1. Introduction

Bovine milk provides a variety of important nutritional benefits for the human body, which may include immunological protection and biologically active substances (Clare & Swaisgood, 2000). Milk and milk products are good sources of vitamin D, calcium, magnesium, and potassium (Ranganathan, Nicklas, Yang, & Berenson, 2005; Weinberg, Berner, & Groves, 2004). However, milk consumption among children and adolescents in the United States has been declining since 1977-1978 (Hayden, Dong, & Carlson, 2013; Sebastian, Goldman, Enns, & LaComb, 2010). For most Americans, their consumption of dairy products is below *The Dietary Guidelines for Americans* (Hayden et al., 2013). Flavored milks are very popular among children and adults due to their desirable taste (Kim, Lopetcharat, & Drake, 2013). Accordingly, flavored milk may provide a good opportunity to help meet dietary guideline for dairy products in the United States (Kim et al., 2013; Nicklas, O'Neil, & Fulgoni, 2013)

Consumers frequently add milk to coffee. Consumers prefer milk-based coffee beverages over water-based ones (Parat-Wilhelms et al., 2005). Milk has a significant impact on the coffee beverage's sensory properties, such as appearance, taste, and smell (Richardson-Harman & Booth, 2006). Further, milk can reduce the bitter taste of coffee (Parat-Wilhelms et al., 2005). Dairy-based iced-coffee is described as ‘‘sweet,’’ ‘‘creamy,’’ and “milky,” whereas water-based coffee is often described with either neutral or negative sensory perceptions, such as ‘‘water-like,’’ ‘‘bitter,’’ and ‘‘bland” (Petit & Sieffermann, 2007).

Coffee flavor can be a positive factor for consumer acceptance of a coffee beverage (Li, Hayes, & Ziegler, 2014). However, increasing coffee flavor intensity by adding more coffee extract will inevitably produce more intense bitterness. Bitterness is generally regarded as having a negative impact on consumer acceptance (e.g. Harwood, Ziegler, & Hayes, 2012; Moskowitz & Gofman, 2007). Therefore, a trade-off has to be made to reach an optimal formulation. This trade-off decision can be made through optimization techniques.

Optimization is an important practice for product developers (Ares, Varela, Rado, & Giménez, 2011; Dutcosky, Grossmann, Silva, & Welsch, 2006) to achieve a competitive status in markets (Stone & Sidel, 2004; Villegas, Tarrega, Carbonell, & Costell, 2010). Due to intense competition in the market, the food industry is increasingly interested in optimization tools and techniques that can be used rapidly and easily to save time and cost of product development. Operationally, optimization can be approached in two distinct ways: by maximizing overall acceptability (e.g., (Deshpande, Chinnan, & McWatters, 2008; Youn & Chung, 2012) or by minimizing dissatisfaction. Recently, Just-About-Right (JAR) scales (Popper & Gibes, 2004; Rothman & Parker, 2009; Xiong & Meullenet, 2006), which seek to minimize dissatisfaction (i.e. “Too Little” and “Too Much”), have gained popularity as an optimization technique because they are rapid and easy to perform.

Using JAR scaling, an attribute is evaluated for its appropriateness relative to an ideal level (Rothman & Parker, 2009; Worch et al., 2010). This hypothetical ideal is designated “Just About Right” or “Just Right.” Accordingly, a participant may indicate an attribute is “Too Little”, “Too Much” or “Just About Right.” Generally, when an attribute is “Too Little” or “Too Much”, it will be optimized by increasing or decreasing the amount of the ingredient that corresponds to that attribute. This technique is useful when a systematic solution (e.g., full formulation design) is not available, or when cost or time is a matter of concern. However, some recommend against replacing traditional experimental design with JAR scaling for product optimization (Stone & Sidel, 2004). JAR scaling is criticized for its practice of combining the measurements of attribute intensity and consumer acceptability into one measurement scale (Moskowitz, Muñoz, & Gacula, 2008). Additionally, JAR scales may suffer from other flaws that interfere with optimization, such as attitudinal biases unrelated to sensory properties or a lack of attribute independence (Rothman & Parker, 2009).

As an alternative to JAR scaling, Ideal scaling measures attribute perceived intensity and subjective ideal intensity separately (Gilbert, Young, Ball, & Murray, 1996; Rothman & Parker, 2009; van Trijp, Punter, Mickartz, & Kruithof, 2007; Worch, Le, Punter, & Pages, 2012b). Unlike JAR scaling, where the ideal level (i.e., “Just About Right” or “Just Right”) is fixed at the central point of the scale, ideal scaling allows a participant to designate his/her hypothetical ideal intensity anywhere on the scale. Similarly, the attributes “Too Little” or “Too Much” can be estimated by the deviation (delta) between perceived intensity and ideal intensity.

Using ideal scaling and JAR scaling, an attribute dysfunction level is measured by the deviation between perceived intensity and one’s ideal intensity.

This deviation from the ideal, i.e., “Too little” or “Too much”, is a measure of dissatisfaction in regard to that specific attribute. The farther the attribute intensity deviates from the ideal level, presumably the worse the product quality would be, and the more a consumer would be dissatisfied. Meanwhile, attributes “Too Little” and “Too Much” also indicate what participants say they prefer in terms of intensity. For example, when the coffee flavor of a coffee-flavored beverage is rated as “Too Little”, it tells product developers that participants believe they would prefer a product with stronger “coffee flavor”. We believe it is important to distinguish between minimizing dissatisfaction, as is done in JAR and ideal scaling, and maximizing liking. Notably, in the Kano model, consumer dissatisfaction is not simply the opposite of satisfaction (Berger et al., 1993; Kano, Seraku, Takahashi, & Tsuji, 1984). Further, disparities in optimal levels for a single attribute obtained from JAR scaling and hedonic scores have been widely reported (Bower & Boyd, 2003; Daillant & Issanchou, 1991; Epler, Chambers IV, & Kemp, 1998; Shepherd, Smith, & Farleigh, 1989; van Trijp et al., 2007; Vickers, 1988). These differences are greater when health-concerned attributes are rated.

In this paper, we hypothesized that overall attribute deltas (measured by ideal scaling or JAR scaling) differed from overall liking in terms of measuring product overall quality. Consequently, optimal formulations would differ when these two parameters were optimized to reach a high product quality. The objective of this study was to investigate optimal formulations obtained by maximizing liking as compared to minimizing attribute deltas (dissatisfaction).

2. Materials and methods

This project included two studies, i.e., study I: product optimization, and study II: optimization validation. In study I, product optimization was conducted under three research conditions that differed in research ballot design. In study II, consumer overall liking and preference for two selected optimal formulations were evaluated separately. The ethics statement and method of product preparation were identical for both studies.

2.1 Ethics Statement

The Penn State Office of Research Protections staff exempted the procedures from IRB review under the wholesome foods exemption in 45 CFR 46.101(b)(6). After the participant signed into the computer in the testing booth, informed consent text was presented on the screen. To proceed with the experiment, participants were required to answer a yes/no question to indicate their consent before pressing “continue”. Participants were compensated in cash for their time.

2.2 Sample formulation and preparation

Samples were formulated using coffee extract (3.0-5.0 wt %; Autocrat Sumatra 1397, Autocrat Natural Ingredients, Lincoln, RI), sucrose (5.0-8.0 wt %), milk (35-55 wt %, 2% fat), and water (35-55 wt %). In the optimization study, samples (n=20) were formulated using a fractional mixture design with four constrained variables (Table 4-1). In the validation study, only two samples were tested for liking and preference. These two samples were formulated using optimal formulations (Table 4-2) obtained from the *liking_III* and *JAR_Delta* models in the optimization study. For both studies, formulation variables accounted for 99.8% of the individual formulations. A constant amount of pectin (0.2 wt %; Grinsted® SY, Dupont Danisco) was added to all the samples.

Table 4-1. Sample formulations (in weight percentage) in study I

Product[1]	Milk (%)	Water (%)	Coffee extract (%)	Sucrose (%)	Solid content (%)[2]
1	35.93	54.89	3.99	4.99	10.02
2	45.24	45.24	4.32	4.99	11.11
3, 19	36.93	54.89	2.99	4.99	9.90
4, 9	53.89	34.93	2.99	7.98	14.75
5	34.93	51.90	4.99	7.98	13.14
6, 18	34.93	54.89	4.99	4.99	10.15
7	44.91	43.41	4.99	6.49	12.73
8	35.93	54.39	2.99	6.49	11.29
10	54.89	34.93	4.99	4.99	12.32
11	44.41	44.41	2.99	7.98	13.71
12, 16	54.39	34.93	3.99	6.49	13.53
13	54.89	36.93	2.99	4.99	11.86
14, 17	34.93	53.89	2.99	7.98	12.68
15	51.90	34.93	4.99	7.98	14.99
20	34.93	52.89	3.99	7.98	12.91

[1]Samples in the same row share the same formulation.

[2]Calculated from the solids content of the ingredients.

Table 4-2. Two optimal formulations (in weight percentage) in study II

Samples[1]	Description[2]	Milk	Water	Coffee extract	Sucrose
3% Coffee	*Liking_III*	54.3	35.8	3.0	6.7
5% Coffee	*JAR_Delta*	44.9	43.4	5.0	6.5

[1] For convenience, the sample created using the optimal formulation obtained by *Liking_III* model was identified as 3% coffee; the sample created using the optimal formulation obtained from *JAR_Delta* model was identified as 5% coffee.

[2] Refers to optimization models that determined corresponding formulations.

For convenience, pectin was mixed with sucrose completely before blending them with water, milk, and coffee extract to make sample batches (Figure 1-1). Batches were heated to 72º C and held 15 seconds to assure that the sucrose was completely dissolved, the pectin dispersed and the product was safe for consumption. The finished samples were kept at refrigeration temperature (~4.0°C) for at least 24 hours before serving. Two ounces of coffee milk were served in 4-oz Solo transparent plastic cups (Solo Cup Company, Urbana, IL).

2.3 Participants

Participants were recruited via email using an existing participant database maintained by the Sensory Evaluation Center at Penn State, or via staff intercepts in public spaces in and around the Food Science Department at Penn

State. To qualify for participation, individuals had to be regular drinkers of coffee or coffee-flavored beverages, and free of food allergies.

In study I, participants (n=388,110 men) were randomly assigned to one of three research conditions (for convenience, they were named Method I, Method II, and Method III). The majority of participants (155) were between 18-27 years old, 72 were 28-37, 56 were 38-47, 75 were 48-57, 26 were 58-67, and only 4 were over 67 years old. The majority were White (n=298, ~77%), 59 identified themselves as Asian or Pacific Islander, 9 as African or African American, and 11 did not report their race. More than 58% of the participants indicated they drank coffee with milk, cream, and/or sugar with each research categories (Table 4-3).

Table 4-3. Frequency (%) of regularly consumed coffee-flavored beverages for participants in optimization study I.

Product	Method I (n=127)	Method II (n=129)	Method III (n=132)
Cappuccino	19.7	20.9	30.3
Latte	32.3	24.0	37.9
Black Coffee	21.3	27.9	25.8
Iced Coffee	33.9	25.6	37.9
Coffee with milk, cream, and/or sugar	60.6	61.2	58.3

Note: This is a “check all that apply” question. So the sums of percentage in a column may exceed 100%.

In study II, participants (n=122) were recruited and randomly assigned into either a liking (acceptance) test or a preference test. Gender distribution in the two research conditions was similar: 44 female (liking) and 42 female (preference). In the liking test (n=61), about 65% were either 18-27 years old (n=15) or 28-37 years old (n=25), 6 were 38-47, 13 were 48-57, 2 were 58-67; the majority were White (n=51, ~83%), 5 identified themselves as Asian or Pacific Islander, 1 as African or African American, 3 as Hispanic/Latino, and 1 as Other. In the preference test (n=61), just under 60% were between 18-27 years old (n=21) or 28-37 years old (n=14), 12 were 38-47, 13 were 48-57, 1 was 58-67. Similarly, the majority were White (n=53, ~87%), 2 identified themselves as Asian or Pacific Islander, 1 as African or African American, 3 as Hispanic/Latino, and 2 as Other. Compared to the optimization study (Table 4-3), more participants (in the percentages) drank cappuccino, latte, black coffee, and iced coffee; however, fewer participants (~35%) indicated they consume coffee with milk, cream, and/or sugar within each research condition (Table 4-4).

Table 4-4. Frequency (%) of regularly consumed coffee-flavored beverages for participants in validation study II

Products	Liking test (n=61)	Preference test (n=61)
Cappuccino	78.6	72.1
Latte	65.5	72.1
Black coffee	65.5	73.8
Iced coffee	59.0	70.4
Coffee with milk, cream, and/or sugar	32.8	37.7

Note: This is a "check all that apply" question. So the sums of percentage in a column may exceed 100%.

2.4 Product testing

Data collection was conducted using Compusense *five*® software (Compusense Inc., Ontario, Canada). The study protocols differed between study I and II.

2.4.1 Study I: Product Optimization

Participants were randomized to 1 of 3 test conditions upon entering test booths. In method I (n=127), only *liking* and attribute intensities were collected. In method II (n=129), participants rated *liking*, attribute intensities, and their ideal attribute intensities on separate, appropriately-worded line scales. In method III (n=132), *liking* was collected, and attribute appropriateness was assessed with Just-About-Right (JAR) line scales.

Liking was assessed using a standard 9-point hedonic scale (1= “Dislike Extremely”, 5 = “Neither Like Nor Dislike”, and 9=“Like Extremely”) (Peryam & Pilgrim, 1957). Attribute intensities, both perceived and ideal, were measured using continuous line scales (0-100); two descriptive anchors were placed at 10% and 90% of these scales, representing low intensity (e.g., “Not At All Sweet”) and high intensity (e.g., “Extremely Sweet”). Just-About-Right (JAR) scales were designed as continuous line scales with three descriptive anchors, low intensity (i.e., “Much Too Weak”) on the left end, “Just About Right” at the middle, and high intensity (i.e., “Much Too Strong”) on the right end. Demographics and consumption behavior for coffee-flavored beverages were collected after all samples had been evaluated.

To minimize sensory fatigue, participants received 4 formulas out of 20 using an incomplete block design. The samples were served in a monadic sequential order, with a two-minute mandatory break between samples. During the break, participants were asked to rinse with room temperature (22°C) filtered water to reduce potential carry-over effects.

2.4.2 Study II: Optimization Validation

In the liking test, the two samples were rated for overall liking using a random complete block design; samples were served in a monadic sequential order. In the preference test, the two samples were served in pairs. Participants were asked to rinse with room temperature (22°C) filtered water between samples to reduce potential carry-over effects.

Liking was assessed using a 9-point Quartermaster hedonic scale (1= “Dislike Extremely”, 5 = “Neither Like Nor Dislike”, and 9=“Like Extremely”) (Peryam & Pilgrim, 1957) (see Appendix D). The preference test was designed as a 2AFC test: a no preference option was not provided (Appendix E). Demographics and consumption behavior for coffee-flavored beverages were also collected after the samples had been evaluated.

2.5 Statistical analyses

In study I, mean *liking* was not significantly influenced by research method (method) (Li et al. 2014), *so liking* data were aggregated and mean liking (*Overall_Liking*) for each sample (n=20) was calculated across all methods and panelists. *Overall_liking* was regressed on formulation variables (coffee, milk, sucrose, and water) to yield an optimal formulation using eChip® software (Wilmington, DE).

Similarly, to calculate optimal formulae for the individual methods, mean *liking* scores were calculated for each sample across all the participants within a research method (Method I, Method II, and Method III). For convenience, these mean liking scores within each method were identified as *Liking_I*, *Liking_II*, and *Liking_III*, respectively. In Method II and Method III, attribute delta (i.e., “Too Little” or “Too Much”) was calculated as the absolute deviation of perceived intensity from ideal intensity in the ideal scaling, or “just about right” level in the

JAR scaling. In ideal scaling, the mean of ideal intensities (n=4) for that participant and that attribute were used as the veridical ideal point for that individual. Using these attribute deltas, we created *Ideal_Delta* and *JAR_Delta* variables to estimate overall product quality within each scaling method. For each sample, *Ideal_Delta* or *JAR_Delta* was estimated as the sum of four attributes deltas (*sweetness, milk flavor, thickness,* and *coffee flavor*) averaged across participants within each scaling method, as follows:

Ideal_Delta, or *JAR_Delta* =

$$\sum(|delta_{sweetness}| + |delta_{coffee\ flavor}| + |delta_{milk\ flavor}| + |delta_{thickness}|)$$

Overall_liking, *Liking_I*, *Liking_II, Liking_III*, *Ideal_Delta* and *JAR_Delta* were fitted as a function of formulation variables (coffee extract, sucrose, milk, and water). To achieve optimal formulations, response variables *Overall_liking*, *Liking_I*, *Liking_II*, and *Liking_III* were maximized or *Ideal_Delta* and *JAR_Delta* were minimized.

In study II, data were analyzed using JMP® version 9.02 (SAS Institute Inc.). Mean *liking* for the two samples were compared using an analysis of variance, where participant was treated as a random effect and sample was a fixed effect. Mean *liking* for each sample was also compared to the corresponding predicted optimal liking values to test the predictive ability of optimization models. Preference data were analyzed using a binomial test to see if one sample was significantly preferred over the other.

3. Results

3.1 Study I: product optimization using *Overall_Liking* (n=388)

The regression model explained 75.8% of the variation in *Overall_liking* (p=0.03). Only the variables of water (p=0.019), sucrose (p=0.006), milk*sucrose (p=0.029), and sucrose^2 (p=0.025) were significant in the final model (Table 4-5). Surprisingly, all coffee-related variables were not significant. Using this prediction model, the optimal formulation for a coffee-flavored dairy beverage was determined as milk = 54.2, water = 35.6, coffee extract = 3.0, and sucrose = 7.0 weight % (Figure 4-1). This optimal beverage is predicted to have a mean *liking* of 6.93 (95% CI of 5.97-7.91), which is close to 7.0 (i.e., “Like Moderately” on a 9-point hedonic scale)

Table 4-5. *Overall_Liking* optimization model (n=388)

Predictor variables	Coefficients	p-value
Intercept	7.06	--
Milk	-0.43	0.3036
Water	**-0.84**	**0.0185**
Coffee	-3.21	0.4993
Sucrose	**10.62**	**0.0061**
Milk*water	-0.61	0.3452
Milk*coffee	-2.62	0.6723
Milk*sucrose	**10.28**	**0.0285**
Water*coffee	3.27	0.5340
*Water*sucrose*	*7.36*	*0.0848*

Coffee*sucrose	5.37	0.8610
$Milk^2$	-0.73	0.2922
$Water^2$	-0.82	0.1447
$Coffee^2$	-22.15	0.8162
$Sucrose^2$	**-116.07**	**0.0246**

Note: p-values in bold indicate terms are significant in the model at α=0.05.

Italicized p-values are significant at α=0.10.

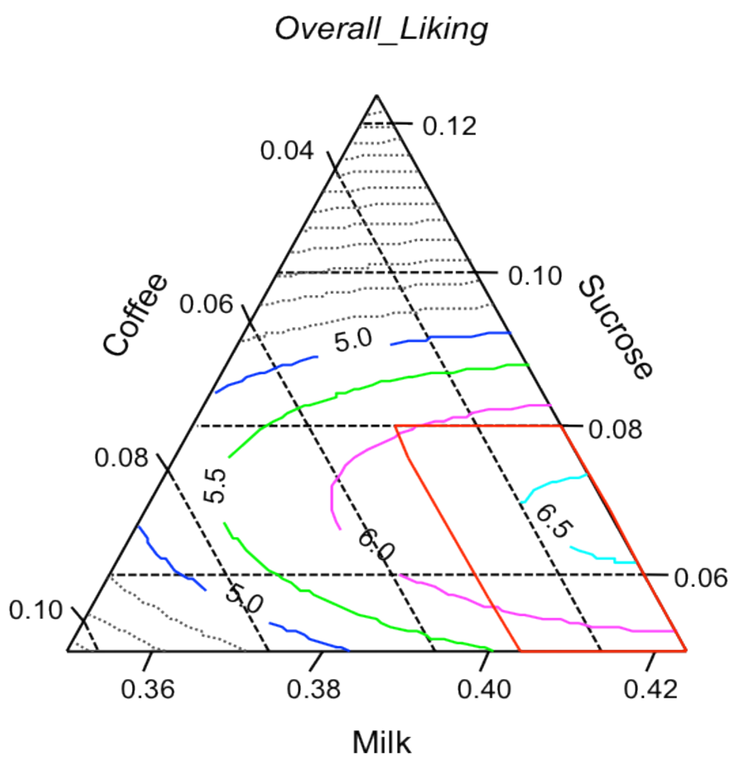

Figure 4-1. Contour plot for product optimization using *Overall_Liking*

Notes: 1. Solid lines in the contour plot indicate that predicted responses were significantly different from each other (α=0.05). 2. Dashed lines refer to predicted responses outside of the observed range of liking. 3. Contour lines are placed at the least significant difference between liking values. 4. The parallelogram defines the experimental space.

3.2 Study I: product optimization using *Liking_I* (n=127)

The regression model explained 77.6% of the variation of *Liking_I* (p = 0.0236). Only the variables water (p=0.0035) and sucrose (p=0.0220) were significant in the final model (Table 4-6). An optimal formulation for coffee milk was determined as milk = 49.2, water = 38.7, coffee extract = 4.2, and sucrose = 7.7 weight % (Figure 4-2). This optimized coffee milk is predicted to have an average *liking* of 7.19 (95% CI of 5.91-8.47), which is close to 7.0 (“Like Moderately”) on a 9-point hedonic scale.

Table 4-6. *Liking_I* optimization model (n=127)

Predictor variables	Coefficients	p-value
Intercept	7.45	--
Milk	-0.24	0.6670
Water	**-1.52**	**0.0035**
Coffee	1.96	0.7573
Sucrose	**11.21**	**0.0220**
Milk*water	0.49	0.5671
Milk*coffee	-3.91	0.6411
Milk*sucrose	8.73	0.1383
Water*coffee	2.53	0.7196

Water*sucrose	-0.19	0.9718
Coffee*sucrose	66.95	0.1276
Milk^2	*-1.71*	*0.0811*
Water^2	-0.58	0.4218
Coffee^2	-98.79	0.4482
Sucrose^2	-91.47	0.1533

Note: p-values in bold indicate terms are significant in the model at α=0.05.

Italicized p-values are significant at a=0.10.

Liking_I

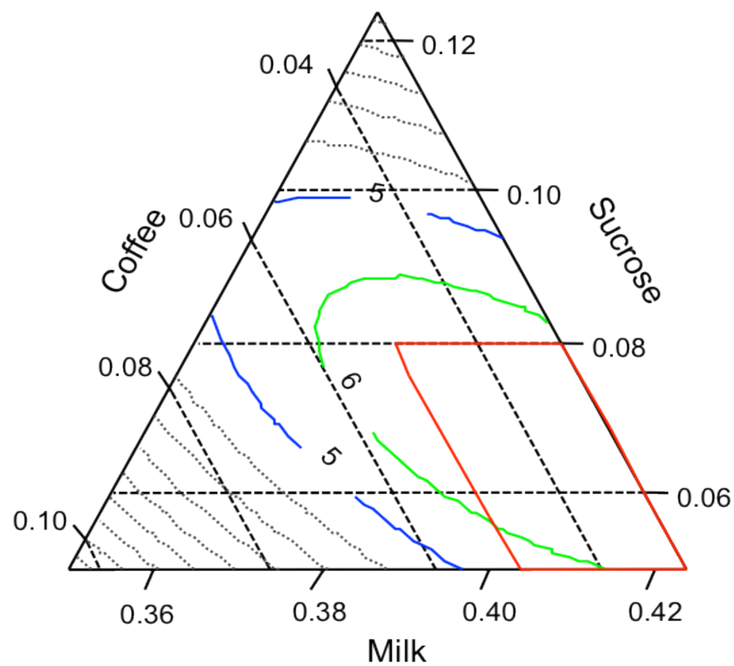

Figure 4-2. Contour plots for product optimization using *Liking_I* (n=127)

Notes: 1. Solid lines in the contour plot indicate that predicted responses were significantly different from each other (α=0.05). 2. Dashed lines refer to predicted responses outside of the observed range of liking. 3. Contour lines are placed at the least significant difference between liking values. 4. The parallelogram defines the experimental space.

3.3 Study I: product optimization using *Liking_II* and *Ideal_Delta* (n=129)

The regression model explained 62.5% of the variation in *Liking_II* (p = 0.1756). None of the terms were significant, although water, sucrose, milk*sucrose, and sucrose^2 terms were all marginal (Table 4-7). An optimal formulation for coffee milk was estimated as milk = 48.3, water = 41.5 coffee extract = 3.4, and sucrose = 6.6 weight % (Figure 4-3, left). This coffee milk formula is predicted to have a mean *liking* of 6.9 (95% CI of 5.8-8.0), which is also close to 7.0 ("Like Moderately") on the 9-point hedonic scale.

Using *Ideal_Delta*, the regression model explained 38.0% of the variation of *Ideal_Delta*, and was not significant (p =0.7115) (Table 4-7). An optimal formulation (in weight percentage) for coffee milk was estimated as milk = 44.9, water = 43.4, coffee extract = 5.0, and sucrose = 6.5 (Figure 4-3, right).

Table 4-7. *Liking_II* and *Ideal_Delta* optimization model (n=129)

Predictor variables	*Liking_II* model		*Ideal_Delta* model	
	Coefficients	p-value	Coefficients	p-value
Intercept	7.31	--	50.12	--
Milk	-0.26	0.5768	14.87	0.2651
Water	*-0.73*	*0.0529*	15.74	0.1261
Coffee	0.63	0.9042	-101.45	0.5001
Sucrose	*6.42*	*0.0919*	-135.57	0.1934

Milk*water	-0.33	0.6418	12.93	0.5212
Milk*coffee	0.09	0.9893	-76.58	0.6965
Milk*sucrose	7.92	0.1094	-42.12	0.7475
Water*coffee	6.49	0.2808	-11.55	0.9441
Water*sucrose	4.35	0.3371	-84.41	0.5041
Coffee*sucrose	16.83	0.6265	-343.93	0.7238
$Milk^2$	-0.97	0.2147	-0.57	0.9784
$Water^2$	-0.91	0.1489	2.39	0.8865
$Coffee^2$	-101.52	0.3525	1363.34	0.6527
$Sucrose^2$	-87.95	0.1044	1075.83	0.4568

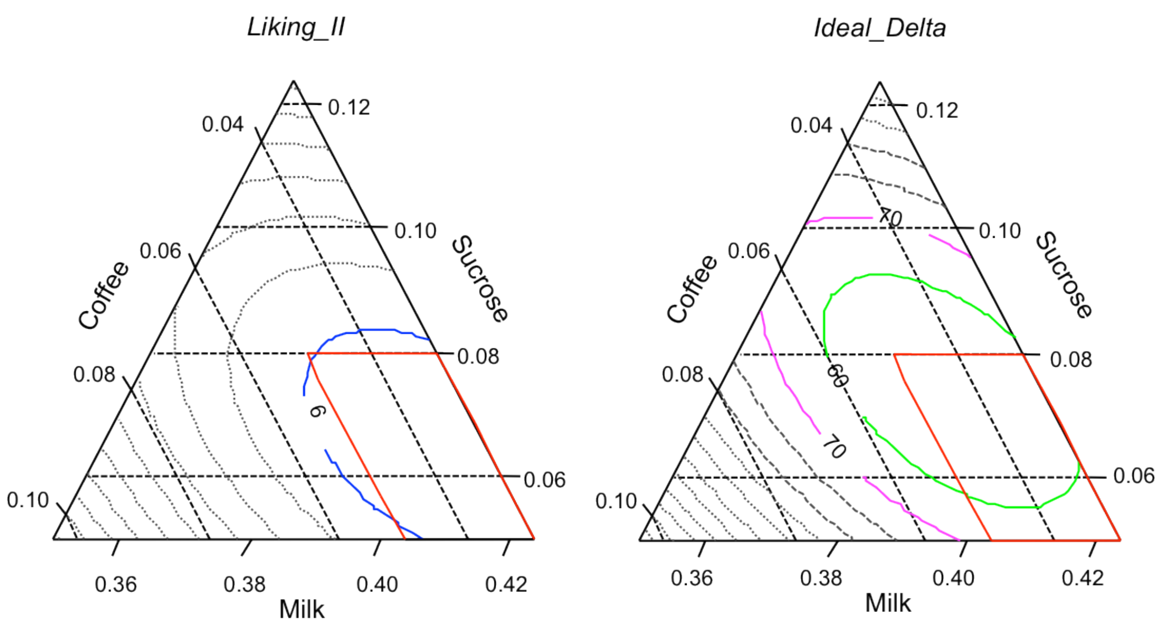

Figure 4-3. Contour plots for product optimization using *Liking_II* and *ideal_Delta* *(n=129)*

Notes: 1. Solid lines in the contour plot indicate that predicted responses were significantly different from each other (α=0.05). 2. Dashed lines refer to predicted responses outside of the observed range of liking or *Ideal_Delta*. 3. Contour lines are placed at the least significant difference between liking or *Ideal_Delta* values. 4. The parallelogram defines the experimental space.

3.4 Study I: product optimization using *Liking_III* and *JAR_Delta* (n=132)

The regression model explained 72.1% of the variation in *Liking_III* (p=0.0581). Sucrose, milk*water, milk*sucrose, water*sucrose and sucrose^2 significantly contributed to variation in *Liking_III* (Table 4-8). An optimal formulation (in weight percentage) for coffee milk was determined as milk = 54.3, water = 35.8, coffee extract = 3.0, and sucrose = 6.7 (Figure 4-4, left). This optimized coffee milk is predicted to have a mean *liking* of 7.1 (95% CI of 5.8-8.5).

For *JAR_Delta*, the regression model explained 71.5% of the variation (p=0.0632). The milk, water, sucrose, milk*water, milk*sucrose, water*sucrose, and sucrose^2 contributed significantly to variation in *JAR_Delta* (Table 4-8). An optimal formulation (in weight percentage) for coffee milk was determined as milk = 44.9, water = 43.4, coffee extract = 5.0, and sucrose = 6.5 (Figure 4-4, right). Notably, this optimal formulation is identical to the one obtained from the *Ideal_Delta* model.

Table 4-8. *Liking_III and JAR_Delta* optimization model (n=132)

Predictor variables	*Liking_III* model		*JAR_Delta* model	
	Coefficients	p-value	Coefficients	p-value
Intercept	6.39	--	30.34	--
Milk	-0.72	0.2113	**22.14**	**0.0330**
Water	-0.31	0.4638	**15.25**	**0.0464**
Coffee	-12.04	0.0820	62.41	0.5585
Sucrose	**14.07**	**0.0071**	**-278.06**	**0.0024**
Milk*water	**-1.91**	**0.0461**	**35.39**	**0.0285**
Milk*coffee	-4.14	0.6244	61.66	0.6591
Milk*sucrose	**13.98**	**0.0286**	**-271.11**	**0.0135**
Water*coffee	-0.18	0.9796	-23.56	0.8409
Water*sucrose	**17.12**	**0.0084**	**-282.25**	**0.0086**
Coffee*sucrose	-62.94	0.1525	628.07	0.3726
$Milk^2$	0.44	0.6333	-4.49	0.7658
$Water^2$	-0.83	0.2626	12.88	0.2932
$Coffee^2$	138.26	0.2992	-1273.33	0.5560
$Sucrose^2$	**-164.88**	**0.0201**	**3213.47**	**0.0087**

Note: p-values in bold indicate terms are significant in the model at α=0.05.

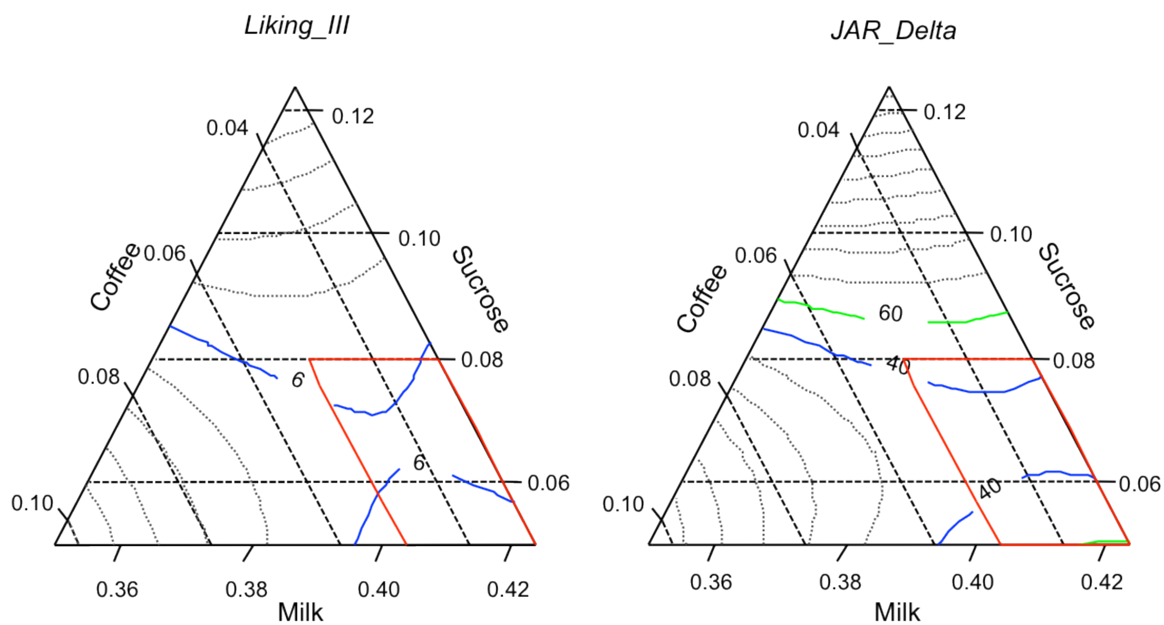

Figure 4-4. Contour plots for product optimization using *Liking_III* and *JAR_Delta* (n=132)

Notes: 1. Solid lines in the contour plot indicate that predicted responses were significantly different from each other (α=0.05). 2. Dashed lines refer to predicted responses outside of the observed range of liking or *JAR_Delta*. 3. Contour lines are placed at the least significant difference between liking or *JAR_Delta* values. 4. The parallelogram defines the experimental space.

3.5 Study II: optimization validation

Two optimal formulations (see Table 4-2) obtained from the *Liking_III* and *JAR_Delta* models were adopted for the optimization validation study. In the validation study, the 3% coffee sample had a mean liking of 7.2, which was significantly higher than the mean liking for the 5% coffee sample (6.4) ($F_{1,60}$=10.93, p=0.0016). In the preference test (n=61), the 3% coffee sample was also significantly preferred when compared to the 5% coffee sample (40 vs 21) ($p=0.0102$). This result was in agreement with the study by Epler et al. (1998), where the participants preferred “the most liked” sweetness level over the “just right” level in a lemonade beverage.

Mean liking (7.2) for the 3% coffee sample was not significantly from the predicted optimal liking (7.1) obtained in the *Liking_III* model ($t=0.231$, $p=0.8181$). Mean liking (6.4) for the 5% coffee sample was not significantly from the predicted optimal liking (6.4) obtained using *Overall_Liking* model and the optimal formulation that was estimated by the *JAR_Delta* model ($t=0.0539$, $p=0.9572$), but it was significantly different from the predicted optimal liking (7.2) obtained in the *Liking_III* model.

4. Discussion

4.1. Design space

The *Overall_Liking* model identified an optimal formulation. Ideally the optimum should not occur at the boundary of the experimental range as it did in this case for coffee (i.e., 3.0%). In hindsight, the range in concentration of coffee extract we selected was too narrow. The interaction between milk and sucrose indicated that the optimal level of sucrose varied as a function of the amount of milk in the beverage. This result was comparable to a previous finding that optimal sucrose levels differed across low and high milk concentration in coffee milk drinks (Moskowitz, 1985). Milky flavor was also affected by sucrose levels in an instant coffee drink (Varela, Beltrán, & Fiszman, 2014). Surprisingly, none of the coffee-related variables were significant in the model. As a coffee-flavored dairy beverage, coffee extract was an important ingredient for these formulations. However, the absence of a significant effect for coffee extract in the optimization model likely reflects an overly narrow range of concentrations (3.0% to 5.0%) that were very close to the optimum. Presumably, a broader range of concentration (e.g. 2 to 6%), or greater deviation from ideal would have revealed a significant effect of coffee extract on *liking*.

4.2 Comparisons of optimization models

Using the same consumer panel, the variance explained in the *Ideal_Delta* model (38%) was much lower than in the *Liking_II* model (63%). The low R-squared in the *Ideal_Delta* model might be due to additional noise introduced by the multiple ratings steps required in ideal scaling (rating of intensity, followed by rating of ideal). An untrained consumer panel can have a high variance in their intensity ratings (Lawless & Heymann, 2010). In contrast, the variance explained by *JAR_Delta* model (72%) was similar to the *Liking_III* model (72.1%). This suggests *JAR_Delta* may be more effective in detecting changes in independent variables compared to *Ideal_Delta*. Similarly, JAR scaling is also reported to be more discriminative to yogurt sucrose levels than the consumption volume (eaten quantity) (Daillant & Issanchou, 1991). JAR scale also was found to be more effective for measuring attribute dysfunctional levels compared to attribute liking (Ares, Barreiro, & GimÉNez, 2009). Our previous study also showed the attribute

“Too Little” and “Too much” obtained by JAR scaling is a better predictor of *liking* compared to those measured by ideal scaling (Li, Ziegler, & Hayes, submitted)

Compared to other optimization models (*Liking_I*, *Liking_II*, *Ideal_Delta*, and *Liking_III*), the *JAR_Delta* prediction model is more sensitive to the impacts of formulation variable because more variables (main factors and interactions) were significant in the model. It has been reported that Just-About-Right scales was more sensitive than sensory (intensity and liking) scales in terms of reflecting the changes in the formulation variables (Moskowitz, 2001). The *JAR_Delta* model not only identified milk, water and sucrose as significant factors, but also found significant interactions of milk*water, milk*sucrose, water*sucrose*,* and the quadratic effect of sucrose. Sucrose can suppress the intensity of coffee flavor and bitterness (Calvino, Garciamedina, & Comettomuniz, 1990). This is a partial masking or suppression that is common in sensory evaluation, i.e., the intensity of a stimulus becomes weaker in a mixture (Lawless, 1979). The significant quadratic effect of sucrose showed a classic inverted “U” shape relationship between stimulus concentration and consumer liking (Hayes & Duffy, 2008; Keast & Hayes, 2011; Moskowitz, 1971; Pfaffmann, 1980).

4.3 Optimization by liking model and dissatisfaction model

The resulting optimal formulations and predicted liking from the different models are summarized in Table 9. In the dissatisfaction models (*Ideal_Delta* and *JAR_Delta*), participants indicated that they desired a coffee-flavored dairy beverage with a higher concentration of coffee extract (5.0%) and lower concentrations of milk (45.0%) and sucrose (6.5%) compared to the optimal formulations determined by the liking models (*Overall_Liking*, *Liking_I*, *Liking_II*, and *Liking_III*). This finding illustrates a disparity between what consumers stated they want and what they liked most. In some cases, consumers tend to ask for more than what they need, for example, more chocolate was requested than what consumers would actually like (Moskowitz, 2001). The *Overall_Liking* model predicted a *liking* of 6.43 using the optimal formulation resulting from the *Ideal_Delta* and *JAR_Delta* models. This predicted *liking* was lower than that predicted from all other models based on *liking*. Both results support our research hypothesis that optimal formulations obtained from attribute delta models (*Ideal_Delta* and *JAR_Delta*) differ from those from *liking* models. These differences might be due not only to distinct measurements of product acceptability defined by two parameters (i.e., *Ideal_Delta/JAR_Delta* and *liking*), but also to some biases with ideal scaling and JAR scaling.

Table 4-9. Optimal formulations and predicted likings

	Overall (n=388)	Method I (n=127)	Method II (n=129)		Method III (n=132)	
Model	*Overall_Liking*	*Liking_I*	*Liking_II*	*Ideal_Delta*	*Liking_III*	*JAR_Delta*
Milk (%)	54.2	49.2	48.3	44.9	54.3	44.9
Water (%)	35.6	38.7	41.5	43.4	35.8	43.4
Coffee extract (%)	3.0	4.2	3.4	5.0	3.0	5.0
Sucrose (%)	7.0	7.7	6.6	6.5	6.7	6.5
Predicted liking	6.9	7.2	6.9	6.4*	7.1	6.4*

Note: *Predicted likings for both *Ideal_Delta* and *JAR_Delta* were estimated using the *Overall_Liking* optimal model.

Measuring attribute “Too Little” or “Too Much” carries an affective tone (van Trijp et al., 2007). In some sense, we believe these measurements also define consumer dissatisfaction level with an attribute. Logically, in this study *Ideal_Delta* and *JAR_Delta* measured consumer overall dissatisfaction level for overall product quality in terms of the performance of four attributes. In contrast, overall liking is a holistic measurement of consumer satisfaction. In the Kano model, consumer satisfaction and dissatisfaction are two different constructs related to consumer acceptance of a product (Berger et al., 1993; Kano et al., 1984). Consequently, factors driving satisfaction might differ from those driving dissatisfaction (Bi, 2012; Li, 2011). Factors affecting liking of an instant coffee drink are not the same as those affecting disliking (Varela et al., 2014). Accordingly, it is not entirely surprising that optimal formulations resulting from the *Ideal_Delta* and *JAR_Delta* models are different from the formulations obtained from the liking model, even with the same individuals.

In both ideal scaling and JAR scaling, attribute deltas were measured in reference to one’s ideal. These deltas indicated one’s demand relative to his/her ideal that reflects his/her belief or desire. In contrast, liking is a holistic parameter that presumably measures the enjoyment derived from a product or service. Liking motivates consumers to use a food product, and to buy and use a health/cosmetic product (Moskowitz, 2002). The difference between attribute delta and liking reflects a dissociation between attitudinal and behavioral factors (e.g. Drewnowski & Moskowitz, 1985). This suggests that what a consumer states he/she would like is not always the same as what he or she actually likes (For example, many American consumers believe they want ‘dark, rich, hearty’ coffee when they do not). Present result agrees with prior reports. van Trijp et al. (2007) also found optimal attribute intensities for a yogurt product estimated by liking ideal point regression model differed from those obtained by the methods of self-reported, JAR-derived or variant-derived (here we call it ideal scaling).

Some evidence suggests the difference between attribute delta and liking models might be greater when attributes are related to health concerns, such as salt and sucrose levels (Drewnowski & Moskowitz, 1985; Epler et al., 1998). The optimal sucrose level for a lemonade beverage was 9.3% using a JAR scale but 10.3% using a liking scale, and participants significantly prefer the beverage with 10.3% sucrose (Epler et al., 1998). An individual on a diet might be more likely to rate “sweetness” in a beverage as “too much” (Popper & Kroll, 2005). Optimal sweetness intensity (6.43) for a yogurt obtained by a liking model was much higher than those optimal intensities obtained from the methods of self-reported (4.20), JAR-derived (4.33) and variant-derived (4.40) (van Trijp et al., 2007). For a salted snack, the self-reported NaCl ideal level (1.5) was much lower than the one predicted by a liking model (5.1). In contrast, optimal levels for spice obtained by two methods (4.0 vs 4.0) were similar, presumably because spice is a relatively neutral attribute as far as health is concerned (Drewnowski & Moskowitz, 1985).

Similar biases might also occur when attributes have positive or negative associations that are independent of their actual influence on liking. Americans believe they want a “rich, hardy, roast” coffee, thus participants might tend to rate coffee flavor as “not strong enough”. However, on the basis of liking, a lower optimal level of coffee extract was predicted. These holistic liking models might have detected the negative impact of bitterness that resulted from adding more coffee extract into the coffee. The bitterness in coffee is generally regarded as negatively affecting consumer acceptance (even though some appropriate amount of bitterness might drive liking) (Harwood et al., 2012; Moskowitz & Gofman, 2007; Popper & Kroll, 2005). As a result, a product with all “just right” attributes might not be the most liked formation (Moskowitz, 2004; Moskowitz, Munoz, & Gacula, 2003).

4.4 Consumer validation study

The results of study II validated the optimization models obtained for Method III. First, 3% coffee sample was created using the optimal formulation obtained by the *Liking_III* model, and this sample was significantly more liked when compared to 5% coffee sample that was developed using the optimal formulation suggested by the *JAR_Delta* model. This finding is consistent to the result that predicted liking (in the *Overall_Liking* model) of the optimal formulation obtained by *JAR_Delta* model was lower than those predicted liking values in the liking models. In addition, this result also matched our expectation since the 3% coffee sample was formulated using the optimal formulation by maximizing liking.

Notably, mean liking ratings for the 3% coffee and 5% coffee formulas were extremely close to the predicted liking values estimated by the optimization models.

Interestingly, the 5% coffee sample was formulated using the dissatisfaction models (*Ideal_Delta* and *JAR_Delta*), where participants stated they would like a coffee-flavored dairy beverage with less sucrose and milk, and more coffee extract (i.e*.,* a product that was less sweet and milky, and stronger coffee flavor); however, this sample was not preferred by the participants. In the dissatisfaction models, participants thought they knew what they wanted, and they would prefer this “to-be” product. But it turns out the participant did not prefer a product that they asked for or designed by themselves. Participants did not know what they wanted, that is, what they ask for is not exactly what they would like and prefer to have (Moskowitz & Gofman, 2007). Therefore, what consumers state they would like may not be a reliable way to determine an optimal formulation. This suggests maximizing liking is a more reliable tool for determining an optimal formulation in terms of consumer liking and preference, in agreement with prior work (Epler et al., 1998).

4.5 Validity of optimization models

It should be kept in mind that *Ideal_Delta* and *JAR_Delta* were created to reflect overall product quality using the performance (deltas) of four attributes (*sweetness, milk flavor, coffee flavor*, and *thickness)*. Food is a complex matrix. So is consumer consumption behavior. When a participant is asked for his/her level of liking for a product, he or she might assess more than the four sensory properties considered here, as other attributes that might arise from the coffee-flavored milk were not measured in this study. Neglected attributes such as bitterness or mouthcoating might be potentially important for consumer acceptability. As a result, *Ideal_Delta* and *JAR_Delta* might be incomplete measurements for overall product quality compared to holistic ratings of overall liking for a product. In addition, the logic of creating *Ideal_Delta* and *JAR_Delta* might be questioned. In creating *Ideal_Delta* and *JAR_Delta*, both deltas for "Too Little" and "Too Much" were weighed equally. However, studies have shown that attributes "Too Little" and "Too Much" can impact on liking differently (Li et al., submitted; Xiong & Meullenet, 2006). As a result, there might have some disparities between *liking*, and *Ideal_Delta* and *JAR_Delta*. Simply, correlation analysis was used to diagnose the agreement between *Liking_II* and *Ideal_Delta*, and between *Liking_III* and *JAR_Delta*. The correlation between *Liking_II* and *Ideal_Delta* (r=-0.51) was weaker than the one between *Liking_III* and *JAR_Delta* (r=-0.87). We do see these disparities. Therefore, the reasonableness of creating *Ideal_Delta* and *JAR_Delta* might be investigated further.

5. Conclusions

Maximizing liking and minimizing attribute deltas yielded different product optimal formulations. First of all, liking and attribute deltas (*Ideal_Delta* and *JAR_Delta*) are two distinct measurements for product overall quality. Attribute delta (i.e., "Too Little" or "Too Much") measures a subject's dissatisfaction level toward attribute performance relative to his/her hypothetical ideal; in contrast, liking is a measurement reflecting one's satisfaction level towards a product or service. Second, compared to attribute deltas (*Ideal_Delta* and *JAR_Delta*), liking is a more holistic parameter for measuring overall product quality. Our data agree with prior work showing that a product with all the directional attributes at ideal performance levels is not the most liked product (Moskowitz, 2001). Third, attribute deltas indicate what a participant wants or desires an attribute to be, and critically, these measurements may be susceptible to attitudinal biases. Thus, asking a consumer panel to design a product would be misleading. Alternatively, maximizing overall liking is a reliable tool for product development. In addition, liking is a main factor driving consumption (Moskowitz, 2002). Therefore, it might be wiser to understand consumer behavior by considering their overall experience with product rather than simply check product attribute performance (Moskowitz & Gofman, 2007).

Although JAR scaling is still very popular in the industry due to its convenience (Rothman & Parker, 2009), JAR scaling and ideal scaling cannot predict appropriate optimal formulas or attribute intensity levels in terms of overall liking or preference. Attribute deltas defined in JAR scaling were more discriminative to detect changes in formulation variables than those measured by ideal scaling. In addition, considering the greater ease of using JAR scaling, we recommend use JAR scaling as a rapid and easy method for product optimization instead of using ideal scaling. JAR scaling is especially meaningful when time and cost are high-priority concerns. However, it should be kept in mind that both JAR scaling and ideal scaling might partially help improve the product quality rather than improve product quality as a whole as a liking scaling can do.

6. Acknowledgments

This project was partially supported by NIH Grant AI094514 to JEH and GRZ. The authors thank Hanna Schuster for preliminary formulations and Maggie Harding for preparing the coffee milk samples. We also thank the Sensory Evaluation Center staff for their assistance with this test.

Chapter 5

Conclusions and Future Work

1. Overall Description

The initial goal for this project was to optimize a coffee-flavored dairy beverage as a line extension to the current chocolate-flavored milk manufactured and sold by the Creamery retail store at Penn State. Benefiting from our lean experiment design that generated a rich dataset, this thesis is able to explore some interesting topics of product optimization in the field of consumer sensory science, including: 1) interpreting consumer preference using psychohedonic and physicohedonic models; 2) comparing Ideal scaling and JAR scaling for attribute diagnosis and product optimization; 3) comparing liking and dissatisfaction model for product optimization. Main conclusions for this study and suggestions for future work are summarized as follows.

2. Conclusions

Conclusion 1. Both the psychohedonic model and the physicohedonic model are useful for understanding consumer insights about products.

In the psychohedonic model, we found that coffee flavor had a positive influence on consumer liking. However, in the physicohedonic model, coffee extract was found to have a negative influence on liking. If only the physicohedonic model had been applied, a product developer would have been directed to decrease coffee extract concentration, in spite of the positive influence of coffee flavor on liking. However, in the psychophysical model, coffee flavor was significantly related linearly to coffee extract. As more coffee extract was added, the coffee flavor became stronger in the coffee-flavored dairy beverage. So simply decreasing coffee extract would not have helped increase consumer liking; it might have even reduced liking. To optimize the beverage formulation, a trade-off decision would have to be considered to balance out the positive and negative sensory properties created by adding more coffee extract.

Conclusion 2. *Sweetness* and *coffee flavor* were two critical sensory attributes impacting consumer acceptability of a coffee-flavored milk.

Consumers liked sweetness in a coffee-flavored milk. Notably "Too Much" *sweetness* had less impact on consumer liking than "Too Little" *sweetness*. Thus, it would be less risky to develop a coffee-flavored dairy beverage that is "too sweet" rather than one that is "not sweet enough". Coffee flavor was a positive factor for consumer liking. However, "too much" coffee flavor had a slightly higher negative impact on liking. This might be due to the fact that increasing coffee flavor by adding more coffee extract into a coffee-flavored milk might produce unpleasant sensory properties, like bitterness.

Conclusion 3. Just-About-Right scaling and ideal scaling are similar but JAR scaling is more efficient.

Even though JAR scaling and ideal scaling differ in defining a subject's ideal level on the scale, both scales provided similar estimates of "Too Little" and

“Too Much” attribute intensities and their impact on liking. However, we did find ideal scaling was less discriminative then JAR scaling. This might be due to noise arising from the multiple rating steps in ideal scaling. Further, considering advantages of practice, time and cost, JAR scaling is more efficient and should be strongly recommended for use by industry.

Conclusion 4. Minimizing dissatisfaction (attribute deltas, i.e., *Ideal_Delta* and *JAR_Delta*) and maximizing consumer liking yielded different product optimal formulation.

The dissatisfaction (attribute delta) model and the liking model produced two distinct optimal formulations. In the *Ideal_Delta* and *JAR_Delta* models, consumers indicated they wanted a coffee-flavored dairy beverage with more coffee extract, less milk and less sucrose when compared to the formulation achieved using the liking models. This difference was because: 1) both parameters (attribute delta and liking) defined two different constructs about consumer acceptability, i.e., disliking and liking, and; 2) some biases coexisted with delta measurement for attributes that had a positive or negative impact on consumer preference.

Conclusion 5. Asking a consumer panel to design a product can be misleading because they do not exactly know what they want.

The dissatisfaction models (*Ideal_Delta* and *JAR_Delta*) were built using attribute deltas. By minimizing the overall dissatisfaction level, we made the product the consumers indicated they wanted. However, we found this formula was less liked and less preferred in a head to head comparison. Therefore, consumers do not exactly know what they want. Instead, maximizing overall liking is a more valid tool for product development.

Conclusion 6. A lean experimental design is useful for product optimization.

In this study, only 20 prototypes were formulated using a fractional factorial mixture design that varied the ratios of sucrose, liquid milk (2%), coffee extract, and water, and only 5 samples were replicated. This is a highly efficient approach. If a full factorial design had been used, there would have been 81 samples (4 factors @ 3 levels, i.e., 3X3X3X3=81). Creating and evaluating all these formulations would have been very expensive and time consuming. Further, an incomplete balance random block design was applied in the consumer testing, so that each participant tasted only 4 samples out of 20. This design avoided consumer sensory fatigue from testing too many samples in one session. Using this design, we also assured that each sample could be tasted by a similar number of consumers. In summary, a lean design was a powerful way to reach such research objectives, where many variables or samples needed to be considered.

3. Suggestions for future works

Suggestion 1. Expand the concentration range for coffee extract to achieve a more appropriate optimal formulation.

In the optimization prediction model that was achieved using the *Overall_Liking* model, none of coffee-related variables were significant. To make a highly liked coffee-flavored milk, we strongly believe that the amount of coffee extract in the beverage is critical to consumer liking. The non-significance factors are likely due to the narrow concentration range of coffee extract that was used. Too narrow a range of coffee extract might result in less difference among samples. So, in the future work, the concentration range of coffee extract should be expanded, such as 2-6% to ensure the design space includes clearly inferior samples.

Further, applying mean liking scores for each sample neglected the differences among individuals. This practice will decrease the difference among samples in terms of liking. To catch these individual differences and improve the optimization, alternative optimization modeling can be conducted on an individual consumer level. However, an incomplete random block design might become inappropriate because individual participants might taste the sample sets differently under this design. Instead, a complete random block design should be applied to avoid any potential biases,

Suggestion 2. Bitterness should be measured to verify our interpretation that bitterness was increased by adding more coffee extract.

In the psychohedonic model, it was found that increasing coffee flavor showed a marginal and positive impact on liking. However, in the physicohenonic model, coffee extract showed a negative albeit nonsignificant impact on liking.

We concluded that this might be because adding more coffee extract increases the bitterness of the coffee-flavored milk. To verify our conclusion, perceived bitterness should be measured, and the relationship between bitterness and coffee extract concentration can be plotted for interpretation.

Suggestion 3. Consumers should be segmented by pre-screening.

Consumers might show different behavior in their product purchases and consumption. These behaviors might show some patterns, such as similar likings toward products, which indicates potential segmentations. In this study, participants were screened using simple screening criteria, i.e., coffee drinkers and free of food allergies. Some people in our study (n=50, out of 388) only drink black coffee. So data might have been skewed when these people were invited to participate in our coffee milk study. Further, we failed to apply appropriate techniques in “demographic question” sections to segment consumers. Rather than use a “check-all-that-apply” for previous consumption of types of coffee beverages, participants might be classified as “black coffee” drinker only, non-black coffee drinker, etc. These segmentations would be helpful for understanding consumer consumption behavior better.

Suggestion 4. Sensory data for samples in the validation study should be evaluated, which offers a better understanding on samples and consumer preference.

In the validation study, participants did not prefer the 5% coffee sample. We concluded that consumers do not know what they want based on the dissatisfaction model, where only four attributes (*sweetness*, *milk flavor*, *coffee flavor* and t*hickness*) were considered. However, there were noticeable differences between the 3% coffee and 5% coffee samples in terms of color and bitterness. These two attributes might have critical impacts on consumer preference and acceptance. Thus, more attributes could be considered and evaluated for interpreting consumer preference over two samples.

Suggestion 5. Ideal scaling and JAR scaling should be compared on individual products.

Typically, ideal scaling and JAR scaling are used for product optimization by evaluating if attributes are either “Too Little” or “Too Much” on individual products. In our study, these two scaling methods were compared using aggregated data, and it concluded two scaling methods were highly similar in terms of “Too Little” and “Too Much” attribute estimates and their effects on liking. The question might arise as to whether these results would be valid for an individual product. For this reason, it would be more meaningful to validate current results on individual products.

Appendix A: Research method I

Product questionnaire

Please evaluate sample %01, a dairy based coffee-flavored beverage. Which statement best describes your OVERALL LIKING?

Dislike It Extremely	Dislike It Very Much	Dislike It Moderately	Dislike It Slightly	Neither Like Nor Dislike	Like It Slightly	Like It Moderately	Like It Very Much	Like It Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

Rate the sweetness of sample %01...

Not At All Sweet — Extremely Sweet

Rate the milk flavor of sample %01...

Not At All Milk Flavored — Extremely Milk Flavored

Rate the coffee flavor of sample %01…

Not At All Coffee Flavored — Extremely Coffee Flavored

Rate the thickness of sample %01...

Extremely Thin — Extremely Thick

Demographic questionnaire

You are...

- ☐ Male
- ☐ Female

To which age group do you belong?

- ☐ 18-27 years old
- ☐ 28-37 years old
- ☐ 38-47 years old
- ☐ 48-57 years old
- ☐ 58-67 years old
- ☐ Over 67 years old

Which best describes your ethnic origin?

- ☐ American Indian
- ☐ Asian/Pacific Islander
- ☐ African American
- ☐ Caucasian
- ☐ Hispanic/Latino
- ☐ Other

Which style of coffee do you drink most often (Please check all that apply)?

- ☐ Cappuccino
- ☐ Latte
- ☐ Black Coffee
- ☐ Iced Coffee
- ☐ Coffee with milk, cream, and/or sucrose

Appendix B: Research method II

Product questionnaire

Please evaluate sample, a dairy based coffee-flavored beverage. Which statement best describes your OVERALL LIKING?

Dislike It Extremely	Dislike It Very Much	Dislike It Moderately	Dislike It Slightly	Neither Like Nor Dislike	Like It Slightly	Like It Moderately	Like It Very Much	Like It Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

Rate the **sweetness** of sample...

Not At All Sweet — Extremely Sweet

Having tasted sample %01, where would you place your **IDEAL sweetness**?

Not At All Sweet — Extremely Sweet

Rate the **milk flavor** of sample...

Not At All Milk Flavored — Extremely Milk Flavored

Having tasted sample %01, where would you place your **IDEAL milk flavor**?

Not At All Milk Flavored — Extremely Milk Flavored

Rate the **coffee flavor** of sample…

Not At All Coffee Flavored — Extremely Coffee Flavored

Having tasted sample %01, where would you place your **IDEAL coffee flavor**?

Not At All Coffee Flavored — Extremely Coffee Flavored

Rate the **thickness** of sample...

Extremely Thin — Extremely Thick

Having tasted sample %01, where would you place **IDEAL thickness**?

Extremely Thin — Extremely Thick

Demographic questionnaire

You are...

- ☐ Male
- ☐ Female

To which age group do you belong?

- ☐ 18-27 years old
- ☐ 28-37 years old
- ☐ 38-47 years old
- ☐ 48-57 years old
- ☐ 58-67 years old
- ☐ Over 67 years old

Which best describes your ethnic origin?

- ☐ American Indian
- ☐ Asian/Pacific Islander
- ☐ African American
- ☐ Caucasian
- ☐ Hispanic/Latino
- ☐ Other

Which style of coffee do you drink most often (Please check all that apply)?

- ☐ Cappuccino
- ☐ Latte
- ☐ Black Coffee
- ☐ Iced Coffee
- ☐ Coffee with milk, cream, and/or sucrose

Appendix C: Research method III

Product questionnaire

Please evaluate sample %01, a dairy based coffee-flavored beverage. Which statement best describes your OVERALL LIKING?

Dislike It Extremely	Dislike It Very Much	Dislike It Moderately	Dislike It Slightly	Neither Like Nor Dislike	Like It Slightly	Like It Moderately	Like It Very Much	Like It Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

Thinking about the **sweetness** of sample %01, is it...?

Much Too Weak	Just About Right	Much Too Strong

Thinking about the **milk flavor** of sample %01, is it...?

Much Too Weak	Just About Right	Much Too Strong

Thinking about the **coffee flavor** of sample %01, is it...?

Much Too Weak	Just About Right	Much Too Strong

Thinking about the **thickness** of sample %01, is it...?

Much Too Thin	Just About Right	Much Too Thick

Demographic questionnaire

You are...

☐ Male

☐ Female

To which age group do you belong?

☐ 18-27 years old

☐ 28-37 years old

☐ 38-47 years old

☐ 48-57 years old

☐ 58-67 years old

☐ Over 67 years old

Which best describes your ethnic origin?

☐ American Indian

☐ Asian/Pacific Islander

☐ African American

☐ Caucasian

☐ Hispanic/Latino

☐ Other

Which style of coffee do you drink most often (Please check all that apply)?

☐ Cappuccino

☐ Latte

☐ Black Coffee

☐ Iced Coffee

☐ Coffee with milk, cream, and/or sucrose

Appendix D. Validation liking study

Please evaluate sample %01, a dairy based coffee-flavored beverage. Which statement best describes your OVERALL LIKING?

Dislike It Extremely	Dislike It Very Much	Dislike It Moderately	Dislike It Slightly	Neither Like Nor Dislike	Like It Slightly	Like It Moderately	Like It Very Much	Like It Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐

You are...

- ☐ Male
- ☐ Female

To which age group do you belong?

- ☐ 18-27 years old
- ☐ 28-37 years old
- ☐ 38-47 years old
- ☐ 48-57 years old
- ☐ 58-67 years old
- ☐ Over 67 years old

Which best describes your ethnic origin?

- ☐ American Indian
- ☐ Asian/Pacific Islander
- ☐ African American
- ☐ Caucasian
- ☐ Hispanic/Latino
- ☐ Other

Which style of coffee do you drink most often (Please check all that apply)?

- ☐ Cappuccino
- ☐ Latte
- ☐ Black Coffee
- ☐ Iced Coffee
- ☐ Coffee with milk, cream, and/or sucrose

Appendix E. Validation preference study

Please taste two coffee-flavored dairy beverages in front of you, from the left to the right. Which one do you prefer?

331	**576**
☐	☐

You are...

- ☐ Male
- ☐ Female

To which age group do you belong?

- ☐ 18-27 years old
- ☐ 28-37 years old
- ☐ 38-47 years old
- ☐ 48-57 years old
- ☐ 58-67 years old
- ☐ Over 67 years old

Which best describes your ethnic origin?

- ☐ American Indian
- ☐ Asian/Pacific Islander
- ☐ African American
- ☐ Caucasian
- ☐ Hispanic/Latino
- ☐ Other

Which style of coffee do you drink most often (Please check all that apply)?

- ☐ Cappuccino
- ☐ Latte
- ☐ Black Coffee
- ☐ Iced Coffee
- ☐ Coffee with milk, cream, and/or sucrose

References

Ares, G., Barreiro, C., & Giménez, A. N. A. (2009). Comparsion of attribute liking and JAR scales to evaluate the adequacy of senosry attributes of milk deserts Journal of Sensory Studies, 24(5), 664-676.

Ares, G., Varela, P., Rado, G., & Gimenez, A. (2011). Identifying ideal products using three different consumer profiling methodologies. Comparison with external preference mapping. Food Quality and Preference, 22(6), 581-591.

Bennett, S. M., L. S. Zhou and J. E. Hayes (2012). "Using milk fat to reduce the irritation and bitter taste of ibuprofen." Chemosensory Perception 5(3-4): 231-236.

Berger, C., Blauth, R., Boger, D., Bolster, C., Burchill, G., DuMouchel, W., . . . Walden, D. (1993). Kano's methods for understanding customer-defined quality. Center for Quality of Management Journal, 2(4), 33.

Bi, J. (2012). A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. Journal of Sensory Studies, 27(2), 87-101.

Boeneke, C. A., McGregor, J. U., & Aryana, K. J. (2007). A dairy-based espresso beverage manufactured using three different coffee bean roasts. Journal of Food Processing and Preservation, 31(1), 1-12.

Booth, D. A., Thompson, A., & Shahedian, B. (1983). A robust, brief measure of an individual's most preferred level of salt in an ordinary foodstuff. Appetite, 4(4), 301-312.

Bower, J. A., & Boyd, R. (2003). Effect of health concern and consumption patterns on measures of sweetness by hedonic and just-about-right scales. Journal of Sensory Studies, 18(3), 235-248.

Brunso, K., & Grunert, K. G. (2007). Consumer attitude measures and food product development. In M. Hal (Ed.), Consumer-led food product development (pp. 197). Boca Raton, FL: CRC Presss LLC.

Calvino, A. M., Garciamedina, M. R., & Comettomuniz, J. E. (1990). Interactions in caffeine-sucrose and coffee-sucrose mixtures - evidence of taste and flavor suppression. Chem Senses, 15(5), 505-519.

Chu, C. A., & Resurreccion, A. V. A. (2004). Optimization of a chocolate peanut spread using response surface methodology (RSM). Journal of Sensory Studies, 19(3), 237-260.

Cines, B. M., & Rozin, P. (1982). Some aspects of the liking for hot coffee and coffee flavor. Appetite, 3(1), 23-34.

Clare, D. A., & Swaisgood, H. E. (2000). Bioactive milk peptides: A prospectus. Journal of Dairy Science, 83(6), 1187-1195.

Costa, A. I. A., & Jongen, W. M. F. (2006). New insights into consumer-led food product development. Trends in Food Science & Technology, 17(8), 457-465.

Cristovam, E., Russell, C., Paterson, A., & Reid, E. (2000). Gender preference in hedonic ratings for espresso and espresso-milk coffees. Food Quality and Preference, 11(6), 437-444.

Daillant, B., & Issanchou, S. (1991). Most preferred level of sugar: rapid measure and consumption test. Journal of Sensory Studies, 6(3), 131-144.

Deshpande, R. P., Chinnan, M. S., & McWatters, K. H. (2008). Optimization of a chocolate-flavored, peanut-soy beverage using response surface methodology (RSM) as applied to consumer acceptability data. Lwt-Food Science and Technology, 41(8), 1485-1492.

Dooley, L., Threlfall, R. T., & Meullenet, J. F. (2012). Optimization of blended wine quality through maximization of consumer liking. Food Quality and Preference, 24(1), 40-47.

Drewnowski, A. (2001). The science and complexity of bitter taste. Nutrition Reviews, 59(6), 163-169.

Drewnowski, A., & Moskowitz, H. R. (1985). Sensory characteristics of foods: new evaluation techniques. The American Journal of Clinical Nutrition, 42(5), 924-931.

Dupas, C. J., Marsset-Baglieri, A. C., Ordonaud, C. S., Ducept, F. M. G., & Maillard, M. N. (2006). Coffee antioxidant properties: Effects of milk addition and processing conditions. Journal of Food Science, 71(3), S253-S258.

Dutcosky, S. D., Grossmann, M. V. E., Silva, R. S. S. F., & Welsch, A. K. (2006). Combined sensory optimization of a prebiotic cereal product using multicomponent mixture experiments. Food Chemistry, 98(4), 630-638.

Earthy, P. J., MacFie, H. J. H., & Hedderley, D. (1997). Effect of question order on sensory perception and preference in central clocation tirals Journal of Sensory Studies, 12(3), 215-237.

Epler, S., Chambers IV, E., & Kemp, K. E. (1998). Hedonic scales are a better predictor than just-about-right scales of optimal sweetness in lemonade. Journal Sensory Studies, 13, 191–197.

Gacula, M. (2008a). Incomplete block designs. In Design and Analysis of Sensory Optimization (10.1002/9780470385012.ch4pp. 35-44): Food & Nutrition Press, Inc.

Gacula, M. (2008b). Product optimization. In Design and Analysis of Sensory Optimization (pp. 105-229): Food & Nutrition Press, Inc.

Gacula, M., Rutenbeck, S., Pollack, L., Resurreccion, A. V. A., & Moskowitz, H. R. (2007). The just-about-right intensity scale: Functional analyses and relation to hedonics. Journal of Sensory Studies, 22(2), 194-211.

Geel, L., Kinnear, M., & de Kock, H. L. (2005). Relating consumer preferences to sensory attributes of instant coffee. Food Quality and Preference, 16(3), 237-244.

Gilbert, J. M., Young, H., Ball, R. D., & Murray, S. H. (1996). Volatile flavor compounds affecting consumer acceptability of kiwifruit Journal of Sensory Studies, 11(3), 247-259.

Goldman, A. (2005). Ideal scaling, an alternative to Just-about-right scales. Paper presented at the 6th Pangborn sensory science symposium, Harrogate, North Yorkshire, UK.

Greenhoff, K., & MacFie, H. J. H. (1999). Preference mapping in practice. . In H. J. H. MacFie & D. M. H. Thomson (Eds.), Measurement of Food Preferences (pp. 137–165). Gaithersburg, MD: Aspen Publishers, Inc.

Harwood, M. L., Ziegler, G. R., & Hayes, J. E. (2012). Rejection thresholds in chocolate milk: evidence for segmentation. Food Quality and Preference, 26(1), 128-133.
Hayden, S., Dong, D., & Carlson, A. (2013). Why are Americans consuming less fluid milk? A look at generational differences in intake frequency, Retrieved 01-28, 2012, from www.ers.usda.gov/publications/err-economic-research-report/err149.aspx
Hayes, J. E., & Duffy, V. B. (2007). Revisiting sugar-fat mixtures: Sweetness and creaminess vary with phenotypic markers of oral sensation. Chemical Senses, 32(3), 225-236.
Hayes, J. E., & Duffy, V. B. (2008). Oral sensory phenotype identifies level of sugar and fat required for maximal liking. Physiology & Behavior, 95(1-2), 77-87.
Hayes, J. E., Sullivan, B. S., & Duffy, V. B. (2010). Explaining variability in sodium intake through oral sensory phenotype, salt sensation and liking. Physiology & Behavior, 100(4), 369-380.
Hayes, J. E., Raines, C. R., DePasquale, D. A., & Cutter, C. N. (2014). Consumer acceptability of high hydrostatic pressure (HHP)-treated ground beef patties. LWT - Food Science and Technology, 56(1), 207-210.
Hoggan, J. (1975). Ideal Profile Technique in new product development with beer. MBAA Technical Quarterly, 12(2), 81-86.
Hough, G., Sanchez, R., Barbieri, T., & Martinez, E. (1997). Sensory optimization of a powdered chocolate milk formula. Food Quality and Preference, 8(3), 213-221.
Johnson, J. R., & Vickers, Z. M. (1988). A hedonic price index for chocolate chip cookies. In D. M. H. Thomson (Ed.), Food Acceptability (pp. 135-153). New York, NY: Elsevier Science Publisher LTD.
Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive quality and must-be quality. The journal of the Japanese society for quality control, 14(2), 39-48.
Keast, R. S. J. (2008). Modification of the bitterness of caffeine. Food Quality and Preference, 19(5), 465-472.
Keast, S. J., R., & Hayes, J. E. (2011). Successful sodium reduction. The world of food ingredients. Retrieved July 19, 2013, from http://www.foodingredientsfirst.com/magazine-digits/Successful-Sodium-Reduction.html
Kim, M. K., Lopetcharat, K., & Drake, M. A. (2013). Influence of packaging information on consumer liking of chocolate milk. Journal of Dairy Science, 96(8), 4843-4856.
Lawless, H. T. (1977). The pleasantness of mixtures in taste and olfaction. Sensory Processes 1(3): 227.
Lawless, H. T. (1979). Evidence for neural inhibition in bittersweet taste mixtures. Journal of Comparative and Physiological Psychology, 93(3), 538-547.
Lawless, H. T. (1986). Sensory interactions in mixtures. Food Technology, 40(11), 69-69.
Lawless, H. T. (1988). Odour description and odour classification revisited. In D. M. H. Thomson (Ed.), Food Acceptability (pp. 27–40). London: Elsevier Applied Science.

Lawless, H. T., & Heymann, H. (2010). Sensory evaluation of food: Principles and Practices: Springer.
Li, B. (2011). Improvements on Just-About-Right (JAR) scales as product optimization tools using Kano modeling concepts. (1500430 M.S.), University of Arkansas, Ann Arbor. Retrieved from http://ezaccess.libraries.psu.edu/login?url=http://search.proquest.com/docview/899718376?accountid=13158
Li, B., Hayes, J. E., & Ziegler, G. R. (2014). Interpreting consumer preferences: physicohedonic and psychohedonic models yield different information in a coffee-flavored dairy beverage. Food Quality and Preference, http://dx.doi.org/10.1016/j.foodqual.2014.03.001(0).
Li, B., Ziegler, G. R., & Hayes, J. E. (submitted). Just-About-Right and ideal scaling provide similar insights into the influence of sensory attributes on liking. Food Quality and Preference.
Lovely, C., & Meullenet, J. F. (2009). Comparison of preference mapping techniques for the optimization of strawberry yogurt. Journal of Sensory Studies, 24(4), 457-478.
Ma, H. L., Pan, Z. L., Li, B. G., Atungulu, G. G., Olson, D. A., Wall, M. M., & McHugh, T. H. (2012). Properties of extruded expandable breadfruit products. Lwt-Food Science and Technology, 46(1), 326-334.
Macfie, H. J., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. Journal of Sensory Studies, 4(2), 129-148.
Mandel, J. (1991). The validation of measurement through interlaboratory studies. Chemometrics and Intelligent Laboratory Systems, 11(2), 109-119.
Mcbride, R. L., & Booth, D. A. (1986). Using classical psychophysics to determine ideal flavor intensity. Journal of Food Technology, 21(6), 775-780.
Meilgaard, M. C., Civille, G. V., & Carr, B. T. (2007). Sensory evaluations techniques. Boca Raton, FL: CRC Press.
Mela, D. J. (1989). A comparison of single and concurrent evaluations of sensory and hedonic attributes. Journal of Food Science 54(4), 1098-1100.
Meullenet, J. F., Lovely, C., Threlfall, R., Morris, J. R., & Striegler, R. K. (2008). An ideal point density plot method for determining an optimal sensory profile for Muscadine grape juice. Food Quality and Preference, 19(2), 210-219.
Mitchell, M., Brunton, N., & Wilkinson, M. (2009). Optimization of the sensory acceptability of a reduced salt model ready meal. Journal of Sensory Studies, 24(1), 133-147.
Modha, H., & Pal, D. (2011). Optimization of Rabadi-like fermented milk beverage using pearl millet. Journal of Food Science and Technology-Mysore, 48(2), 190-196.
Moskowitz, H. R. (1971). The sweetness and pleasantness of sugars. The American journal of psychology, 84(3), 387-405.
Moskowitz, H. R. (1985). New Driections for Product Testing and Sensory Analysis of Food. Westport, Connecticut: Food & Nutrition Press, Inc.
Moskowitz, H. R. (2001). Sensory directionals for pizza: A deeper analysis. Journal of Sensory Studies, 16(6), 583-600.

Moskowitz, H. R. (2002). Sensory drivers of liking and sensory preference segmentation chemistry of taste (Vol. 825, pp. 214-226): American Chemical Society.

Moskowitz, H. R. (2004). Just about right (JAR) directionality and the wandering sensory unit. In data analysis workshop: getting most out of just-about-right data. . Food Quality and Preference, , 15, 891–899.(891–899).

Moskowitz, H. R., & Gofman, A. (2007). Maxwell house's calculus of coffee. In H. R. Moskowitz & A. Gofman (Eds.), Selling Blue Elephants: how to make great products that people want before they even know they want them (pp. 27-46). Upper Saddle River, New Jersey: Wharton School Publishing.

Moskowitz, H. R., Munoz, A. M., & Gacula, M. C. (2003). Viewpoints and Controversies in Sensory Science and Consumer Product Testing. Trumbull, Connecticut: Food & Nutrition Press.

Moskowitz, H. R., Muñoz, A. M., & Gacula, M. C. (2008). Hedonics, Just-About-Right, purchase and other scales. In Consumer Tests Viewpoints and Controversies in Sensory Science and Consumer Product Testing (10.1002/9780470385128.ch9pp. 145-172): Food & Nutrition Press, Inc.

Nicklas, T. A., O'Neil, C. E., & Fulgoni, V. L. (2013). The nutritional role of flavored and white milk in the diets of children. Journal of School Health, 83(10), 728-733.

Pangborn, R. M. (1982). Influence of water composition, extraction procedures, and holding time and temperature on quality of coffee beverage. Lebensmittel-Wissenschaft & Technologie, 15(3), 161-168.

Pangborn, R. M., Guinard, J. X., & Meiselman, H. L. (1989). Evaluation of bitterness of caffeine in hot chocolate drink by category, graphic, and ratio scaling. Journal of Sensory Studies, 4(1), 31-53.

Parat-Wilhelms, M., Denker, M., Borcherding, K., Hoffmann, W., Luger, A., & Steinhart, H. (2005). Influence of defined milk products on the flavour of white coffee beverages using static headspace gas chromatography-mass spectrometry/olfactometry and sensory analysis. European Food Research and Technology, 221(3-4), 265-273.

Peryam, D. R., & Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. Food Technology, 11(9), A9-A14.

Petit, C., & Sieffermann, J. M. (2007). Testing consumer preferences for iced-coffee: Does the drinking environment have any influence? Food Quality and Preference, 18(1), 161-172.

Pfaffmann, C. (1980). Wundt's schema of sensory affect in the light of research on gustatory preferences. Psychological Research, 42(1-2), 165-174.

Popper, R., & Gibes, K. (2004). Workshop summary: Data analysis workshop: getting the most out of just-about-right data - Abstracts. Food Quality and Preference, 15(7-8), 891-899.

Popper, R., & Kroll, R. D. (2005). Just-About-Right scales in consumer research. Chem Sense, 7(3), 3-6.

Popper, R., Rosenstock, W., Schraidt, M., & Kroll, B. J. (2004). The effect of attribute questions on overall liking ratings. Food Quality and Preference, 15(7-8), 853-858.

Ranganathan, R., Nicklas, T. A., Yang, S.-J., & Berenson, G. S. (2005). The nutritional impact of dairy product consumption on dietary intakes of adults (1995–1996): The Bogalusa heart study. Journal of the American Dietetic Association, 105(9), 1391-1400.
Richardson-Harman, N. J., & Booth, D. A. (2006). Do you like the sight or the feel of milk in coffee? Ecology and effortful attention in differential acuity and preference for sensed effects of milk substitute in vended coffee. Appetite, 46(2), 130-136.
Richardson-Harman, N. J., Stevens, R., Walker, S., Gamble, J., Miller, M., Wong, M., & McPherson, A. (2000). Mapping consumer perceptions of creaminess and liking for liquid dairy products. Food Quality and Preference, 11(3), 239-246.
Rossi, F. (2001). Assessing sensory panelist performance using repeatability and reproducibility measures. Food Quality and Preference, 12(5-7), 467-479.
Rothman, L., & Parker, M. J. (2009). Just-About-Right (JAR) scales: design, usage, benefits and risks: American Society for Testing & Materials.
Sánchez-González, I., Jiménez-Escrig, A., & Saura-Calixto, F. (2005). In vitro antioxidant activity of coffees brewed using different procedures (Italian, espresso and filter). Food Chemistry, 90(1–2), 133-139.
Schiffman, S. S., Booth, B. J., Losee, M. L., Pecore, S. D., & Warwick, Z. S. (1995). Bitterness of sweeteners as a function of concentration. Brain Research Bulletin, 36(5), 505-513.
Sebastian, R. S., Goldman, J. D., Enns, C. W., & LaComb, R. P. (2010). Fluid milk consumption in the United States. What we eat in America, NHANES 2005-2006 Retrieved 01-28, 2014, from http://www.ars.usda.gov/SP2UserFiles/Place/12355000/pdf/DBrief/3_milk_consumption_0506.pdf
Shepherd, R., Smith, K., & Farleigh, C. A. (1989). The relationship between intensity, hedonic and relative-to-ideal ratings. Food Quality and Preference, 1(2), 75-80.
Schutz, H. G. (1983). Multiple-regression approach to optimization. Food Technology, 37(11), 46-62.
Stone, H., & Sidel, J. L. (2004). Sensory Evaluation Practices: San Diego, California: Elsevier Academic Press.
USDA. (2013). Milk production. Retrieved from http://www.usda.gov/nass/PUBS/TODAYRPT/mkpr1113.pdf
van Trijp, H. C. M., Punter, P. H., Mickartz, F., & Kruithof, L. (2007). The quest for the ideal product: Comparing different methods and approaches. Food Quality and Preference, 18(5), 729-740.
Varela, P., Beltrán, J., & Fiszman, S. (2014). An alternative way to uncover drivers of coffee liking: Preference mapping based on consumers' preference ranking and open comments. Food Quality and Preference, 32, Part B(0), 152-159.
Vickers, Z. (1988). Sensory specific satiefy in lemonade using a just right scale for sweetness. Journal of Sensory Studies, 3(1), 1-8.
Vickers, Z., Holton, E., & Wang, J. (1998). Effect of yogurt sweetness on sensory specific satiety. Journal of Sensory Studies, 13(4), 377-388.

Vickers, Z., Holton, E., & Wang, J. (2001). Effect of ideal–relative sweetness on yogurt consumption. Food Quality and Preference, 12(8), 521-526.

Villegas, B., Tarrega, A., Carbonell, I., & Costell, E. (2010). Optimising acceptability of new prebiotic low-fat milk beverages. Food Quality and Preference, 21(2), 234-242.

Vickers, Z. M., Christensen, C. M., Fahrenholtz, S. K., & Gengler, I. M. (1993). Effect of questionaire design and the number of samples tasted on hedonic ratings Journal of Sensory Studies, 8(3), 189-200.

Villegas, B., Tarrega, A., Carbonell, I., & Costell, E. (2010). Optimising acceptability of new prebiotic low-fat milk beverages. Food Quality and Preference, 21(2), 234-242.

Wadhwani, R., & McMahon, D. J. (2012). Color of low-fat cheese influences flavor perception and consumer liking. Journal of Dairy Science, 95(5), 2336-2346.

Weinberg, L. G., Berner, L. A., & Groves, J. E. (2004). Nutrient contributions of dairy foods in the United States, continuing survey of food intakes by individuals, 1994–1996, 1998. Journal of the American Dietetic Association*, 104*(6), 895-902.

Worch, T., Dooley, L., Meullenet, J. F., & Punter, P. H. (2010). Comparison of PLS dummy variables and Fishbone method to determine optimal product characteristics from ideal profiles. Food Quality and Preference, 21(8), 1077-1087.

Worch, T., & Ennis, J. M. (2013). Investigating the single ideal assumption using Ideal Profile Method. Food Quality and Preference, 29(1), 40-47.

Worch, T., Le, S., & Punter, P. (2010). How reliable are the consumers? Comparison of sensory profiles from consumers and experts. Food Quality and Preference, 21(3), 309-318.

Worch, T., Le, S., Punter, P., & Pages, J. (2012a). Assessment of the consistency of ideal profiles according to non-ideal data for IPM. Food Quality and Preference, 24(1), 99-110.

Worch, T., Le, S., Punter, P., & Pages, J. (2012b). Extension of the consistency of the data obtained with the Ideal Profile Method: Would the ideal products be more liked than the tested products? Food Quality and Preference, 26(1), 74-80.

Worch, T., Le, S., Punter, P., & Pages, J. (2013). Ideal Profile Method (IPM): The ins and outs. Food Quality and Preference, 28(1), 45-59.

Xiong, R., & Meullenet, J. F. (2006). A PLS dummy variable approach to assess the impact of jar attributes on liking. Food Quality and Preference, 17(3-4), 188-198.

Youn, K. S., & Chung, H. S. (2012). Optimization of the roasting temperature and time for preparation of coffee-like maize beverage using the response surface methodology. Lwt-Food Science and Technology, 46(1), 305-310.