

Published in final edited form as:

Food Qual Prefer. 2014 October 1; 37: 71–78. doi:10.1016/j.foodqual.2014.04.019.

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

Bangde Li^{1,2}, John E. Hayes^{1,2}, and Gregory R. Ziegler^{2,*}

¹Sensory Evaluation Center, College of Agricultural Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

²Department of Food Science, College of Agricultural Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

Abstract

Just-about-right (JAR) scaling is 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 these two constructs 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". 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 fractional, constrained-mixture design that varied the ratio of water, milk, coffee extract, and sucrose. Participants tasted 4 of 20 prototypes that were served in a monadic sequential order using a balanced incomplete block design. Data reported here are for participants randomly assigned to one of two research conditions: ideal scaling (n=129) or JAR scaling (n=132). For both conditions, participants rated overall *liking* using a 9-point 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 evaluated using the standard deviation of their ideal ratings (n=4). All data from a participant were eliminated from further analyses when his/her standard deviation of the ideal ratings for any of the four rated attributes was identified as a statistical outlier. This resulted in the elimination of 15 of 129 (12 %) of participants in the ideal scaling group. Multiple linear regression was employed to model *liking* as a function of "Too Little" or "Too Much" attribute intensities.

© 2014 Elsevier Ltd. All rights reserved.

*Corresponding Author: Dr. Gregory R. Ziegler, Department of Food Science, Pennsylvania State University, 341 Food Science Building, University Park, PA, 16802, USA, grz1@psu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Mean ideal ratings (averaged across participants) for all four attributes were significantly different from the central point of the scale (i.e., 50). However, *Coffee flavor* was the only attribute for which the mean ideal rating (57.2) fell outside the central 10% (45.0–55.0). Even so, the magnitude of “Too Little” and “Too Much” was not affected by the scaling method. 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 affecting liking.

Keywords

Consumer behavior; scaling bias; product development; milk; coffee; formulation

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). 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 reported to be an easy 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 cost or time is a 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, Smith, & Farleigh, 1989). “Just About Right” or “Just Right” is assumed to be a participant’s ideal level (van Trijp, Punter, Mickartz, & Kruithof, 2007). Using JAR scaling, an attribute is evaluated for its performance (appropriateness) relative to this ideal level (Rothman & Parker, 2009; Worch, Dooley, Meullenet, & Punter, 2010). Attribute performance could be “Too Little”, “Too Much” or “Just About Right”. Generally, “Too Little” or “Too Much” attribute intensity is estimated by the deviation of the participant’s scale rating from the center point of the scale. The intensity of an attribute can be increased if it is perceived as “Too Little,” or decreased if it is perceived as “Too Much”. For this reason, the JAR scale is recognized as a directional tool (Moskowitz, 2001).

JAR scaling combines the measurements of attribute intensity and consumer acceptability (Moskowitz, Muñoz, & Gacula, 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 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). Furthermore, studies find optimal formulations achieved by JAR scaling differ from those predicted by hedonic scores (Epler, Chambers, & Kemp, 1988; Shepherd et al., 1989).

JAR scales may incorporate some unique biases. JAR ratings may be influenced by cognitive factors in addition to perception (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 to eat meat may always rate the meat topping on a pizza “not enough”.

Alternatively, ideal scaling separates the measurements of attribute intensity and acceptability using two identical scales (Gilbert, Young, Ball, & Murray, 1996; Rothman & Parker, 2009; van Trijp et al., 2007; Worch, Le, Punter, & Pages, 2012a). In ideal scaling, acceptability is presumably maximized at the ideal intensity level. Both JAR scaling and ideal scaling implicitly assume a participant has an “ideal” (level) for a specific attribute, which may not be valid if the individual is truly indifferent to changes in that attribute. Moreover, these two methods differ in where the ideal level is assumed to lie. 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 participants’ ideal intensities differ from the central point of the scale, which consequently may influence the measurement of “Too Little” and “Too Much,” and their effect 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 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 study, only the data from research conditions that applied ideal scaling and JAR scaling are discussed. 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 informed consent and were compensated for their time.

2.1 Participants

A total of 261 participants (70 male, 191 female) were invited and finished the product evaluation 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 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 ethnicity.

2.2 Sample formulation and preparation

Using eChip[®] software (Wilmington, DE), 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 composition of each formula is shown in Table 2. Pectin solutions were first prepared by blending pectin into the water. Coffee extract, milk, and sucrose were added to pectin solutions. Batches were heated to 72 °C to assure that the sucrose was completely dissolved, the pectin dispersed, and the product was safe for human testing. The finished prototypes were stored at refrigeration temperature (~4°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 one of two research conditions upon entering the test booths. To minimize fatigue, a balanced incomplete block design (Gacula, 2008) was applied to alleviate carryover effects; accordingly, 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 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 center, 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 sample evaluations.

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

2.4 Data analysis

Data analyses were carried out using JMP® (v9.02, SAS Institute Inc.). Significance criterion were set to $\alpha=0.05$.

“Too Little” and “Too Much” refer to perceived intensity rated below or above either the ideal intensity on the ideal scale or the “Just About Right” point on the JAR scale, respectively. “Too Little” and “Too Much” were calculated as the distance between perceived intensity 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 as any standard deviation exceeding 1.5 times the interquartile range. All data from an individual participant were eliminated from further analyses when the standard deviation of any attribute’s ideal ratings (*sweetness, milk flavor, thickness, and coffee flavor*) for that individual was identified as an outlier.

With the statistical outliers 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* were 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 used as the ideal intensity level for the calculation of “Too Little” and “Too Much” for that attribute and 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 1). Except for one participant (ID=50) who precisely indicated his/her ideal level for each attribute (i.e., standard deviations were “zeros”), participants showed variance in their ideal ratings for all the attributes. Several individuals were identified as outliers (standard deviations exceeded 1.5 times IQR), and as a result, data from 15 out of 129 participants were excluded from further analyses.

The stability of ideal intensity ratings was investigated by evaluating the effect of product using ANOVA (Worch & Ennis, 2013). All the ANOVA models had R-squares (adjusted)

that were greater than 85% (Table 3). Product showed a marginal effect on ideal ratings of *sweetness*. Product did not show a significant effect on ideal ratings of other attributes. Serving order, i.e. 1st, 2nd, 3rd, or 4th, significantly influenced ideal ratings for *coffee flavor*. However, 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=56.3, and 4th=56.4).

3.2 Distribution characteristics of ideal intensity ratings

For the ideal scaling condition, participants were allowed to rate their ideal intensities freely on the scale. The distributions of mean ideal intensities for attributes are illustrated in Figure 2. 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 standard deviations greater than 10. Overall means of ideal intensities were calculated and compared to the central point of the scale (i.e. 50); means were 57.2 (p<0.0001) for *coffee flavor*, 47.4 (p=0.0477) for *sweetness*, 47.5 (p=0.0567) for *milk flavor*, and 45.2 (p<0.0001) for *thickness*.

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. Since “Too Little” and “Too Much” were defined as the deviations of perceived intensities from ideal intensities, this resulting asymmetry of ideal rating may influence measurement of “Too Little” and “Too Much” Values of “Too Little” and “Too Much” for all four rated attributes between two scaling were compared (Table 4).

None of the interaction terms (product by method) were significant (p>0.05). 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).

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 tolerance for “Too Little” and “Too Much”. 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 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 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, “Too Little” *sweetness* showed the highest negative 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.

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 et al., 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 a panelist’s ratings (Mandel, 1991; Meilgaard, Civille, & Carr, 2007; Rossi, 2001). Here, with 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 is no specific rule for evaluating the stability of ideal ratings using its standard deviation and scaling range in the literature. However, it was reported that even a well-trained descriptive panel will have standard deviations around 10% of scale range for attribute intensity ratings of a stimulus (Lawless, 1988). Consumer panels generally perform even 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 ($p=0.0493$) showed a marginally significant effect on ideal *sweetness* only. This effect was due to one sample (identified as sample #6, Table 2). Since its replicate (identified as sample #18, table 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 the data related to this sample (sample #6) was eliminated. Ideal ratings across serving orders were also compared. *Coffee flavor* was the only attribute whose ideal ratings significantly differed across serving order ($p=0.0037$). Interestingly, the mean values of ideal *coffee flavor* seemed to decrease with order (1st=58.7, 2nd=57.1, 3rd=56.3, and 4th=56.4), i.e. the more coffee milk a participant tasted, the less intense their desired ideal became. However, this slight difference may not be of practical importance, given how small the absolute changes in ratings were. Generally, the ideal ratings of panel performance were stable.

4.2 “Too Little” and “Too Much;” JAR scaling vis-à-vis ideal scaling

JAR scaling and ideal scaling differ in how they define a participant’s ideal level. In ideal scaling, participants used nearly the entire range of the scale for their ideal ratings. In contrast, constraining a subject’s ideal level to 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 attributes “Too Little” and “Too Much”. JAR scaling and ideal scaling appear to be highly comparable in measuring attribute intensities relative to an implicit ideal.

Some notable differences between the two scaling methods were observed. All the “Too Little” and “Too Much” scores significantly influenced *liking* in JAR scaling, while for ideal scaling, “Too Much” *milk flavor*, and both “Too Little” and “Too Much” *thickness* did not show significant impact on *liking* (Figure 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 can better predict liking when compared to those achieved by the 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 attributes “Too Little” and “Too Much” on liking

With both the JAR scaling and ideal scaling, the attributes “Too Little” and “Too Much” affect *liking* asymmetrically (Figure 4). Participants showed different tolerance levels for deviation from their ideals depending whether they were “Too Little” or “Too Much”. This result is similar to previous studies (Moskowitz, 2001; Xiong & Meullenet, 2006). Both scaling methods agreed *sweetness* and *coffee flavor* were more important factors to consumer *liking* when compared to *milk flavor* and *thickness*. This finding matches our expectation about the importance of *sweetness* and *coffee flavor* for *liking* of 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 5). The classic inverted U shaped relationship between liking and attribute intensity reported in the literature (Keast & Hayes, 2011; Moskowitz, 1971; Pfaffmann, 1980) may really be an “L”. In this case, *sweetness* of a coffee-flavored dairy beverage seemed to fit this pattern well.

Compared to “Too Much”, “Too Little” *sweetness* showed a higher impact on *liking*. This means consumers prefer 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 meaningful for product development, as it is less risky to make a coffee-flavored dairy beverage “Too Sweet” rather than “Not Sweet Enough”. Seemingly, 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 negative impact on *liking* than “Too Little” *coffee flavor*.

In addition to increasing *coffee flavor*, adding more coffee extract also increased bitterness, though we did not ask panelists to rate this attribute. Bitterness is generally regarded as a negative factor to consumer *liking*, and it is possible that dumping of their ratings of bitterness into *coffee flavor* caused the decrease in liking above the ideal *coffee flavor* level. This result is compatible with our previous findings on psychohedonic and physiohedonic models (Li, Hayes & Ziegler, 2014).

5. Conclusion

Sweetness and *coffee flavor* were two critical sensory attributes that directly impact consumer acceptability for a coffee-flavored dairy beverage. “Too Much” *sweetness* had less negative affect on consumer liking than “Too Little” *sweetness*. Thus, it will be less risky for a product developer to have a “too sweet” coffee-flavored dairy beverage than one that “not sweet enough”. Coffee extract is a complex ingredient. Adding more coffee extract into a coffee-flavored beverage would 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 well designed to balance positive and negative sensory perceptions.

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 development (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 more efficient for product evaluation due to decreased panelist demand, and fewer data analysis steps.

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 Feeny for her critical comments on this manuscript.

References

- Ares G, Barreiro C, Giménez ANA. Comparison of attribute liking and JAR scales to evaluate the adequacy of sensory attributes of milk deserts. *Journal of Sensory Studies*. 2009; 24(5):664–676.
- Booth DA, Thompson A, Shahedian B. A robust, brief measure of an individual’s most preferred level of salt in an ordinary foodstuff. *Appetite*. 1983; 4(4):301–312. [PubMed: 6670860]
- Epler S, Chambers E IV, Kemp KE. Hedonic scales are a better predictor than just-about-right scales of optimal sweetness in lemonade. *Journal Sensory Studies*. 1988; 13:191–197.
- Gacula, M. Design and Analysis of Sensory Optimization. Food & Nutrition Press, Inc; 2008. Incomplete Block Designs; p. 35-44.
- Gacula M, Rutenbeck S, Pollack L, Resurreccion AVA, Moskowitz HR. The just-about-right intensity scale: Functional analyses and relation to hedonics. *Journal of Sensory Studies*. 2007; 22(2):194–211.

- Gilbert JM, Young H, Ball RD, Murray SH. Volatile flavor compounds affecting consumer acceptability of kiwifruit. *Journal of Sensory Studies*. 1996; 11(3):247–259.
- Goldman, A. Ideal scaling, an alternative to Just-about-right scales. Paper presented at the 6th Pangborn sensory science symposium; Harrogate, North Yorkshire, UK. 2005.
- Hoggan J. Ideal profile technique in new product development with beer. *MBAA Technical Quarterly*. 1975; 12(2):81–86.
- Keast, SJR.; Hayes, JE. Successful Sodium Reduction. *The world of food ingredients*. 2011. Retrieved July 19, 2013, from <http://www.foodingredientsfirst.com/magazine-digits/Successful-Sodium-Reduction.html>
- Lawless, HT. Odour description and odour classification revisited. In: Thomson, DMH., editor. *Food Acceptability*. London: Elsevier Applied Science; 1988. p. 27-40.
- Lawless, HT.; Heymann, H. *Sensory Evaluation of Food: Principles and Practices*. Springer; 2010.
- Li, B. ProQuest Dissertations & Theses A&I database. University of Arkansas; Ann Arbor: 2011. Improvements on Just-About-Right (JAR) scales as product optimization tools using Kano modeling concepts. (1500430 M.S.).
- Li B, Hayes JE, Ziegler GR. Interpreting consumer preferences: Psychohedonic and psychohedonic models yield different information in a coffee-flavored dairy beverage. *Food Quality and Preference*. 2014; 36(0):27–32. [PubMed: 25024507]
- Lovely C, Meullenet JF. Comparison of Preference Mapping Techniques for the Optimization of Strawberry Yogurt. *Journal of Sensory Studies*. 2009; 24(4):457–478.
- Macfie HJ, Bratchell N, Greenhoff K, Vallis LV. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies*. 1989; 4(2):129–148.
- Mandel J. The validation of measurement through interlaboratory studies. *Chemometrics and Intelligent Laboratory Systems*. 1991; 11(2):109–119.
- Mcbride RL, Booth DA. Using classical psychophysics to determine ideal flavor intensity. *Journal of Food Technology*. 1986; 21(6):775–780.
- Meilgaard, MC.; Civille, GV.; Carr, BT. *Sensory Evaluations Techniques*. Boca Raton, FL: CRC Press; 2007.
- Moskowitz HR. The sweetness and pleasantness of sucroses. *The American Journal of Psychology*. 1971; 84(3):387–405. [PubMed: 5142585]
- Moskowitz HR. Sensory directionals for pizza: A deeper analysis. *Journal of Sensory Studies*. 2001; 16(6):583–600.
- Moskowitz, HR.; Muñoz, AM.; Gacula, MC. Viewpoints and Controversies in Sensory Science and Consumer Product Testing. Food & Nutrition Press, Inc; 2008. Hedonics, Just-About-Right, Purchase and Other Scales in Consumer Tests; p. 145-172.
- Pangborn RM, Guinard JX, Meiselman HL. Evaluation of bitterness of caffeine in hot chocolate drink by category, graphic, and ratio scaling. *Journal of Sensory Studies*. 1989; 4(1):31–53.
- Peryam DR, Pilgrim FJ. Hedonic scale method of measuring food preferences. *Food Technology*. 1957; 11(9):A9–A14.
- Pfaffmann C. Wundt's schema of sensory affect in the light of research on gustatory preferences. *Psychological Research*. 1980; 42(1–2):165–174. [PubMed: 7394125]
- Popper R, Gibes K. Workshop summary: Data analysis workshop: getting the most out of just-about-right data - Abstracts. *Food Quality and Preference*. 2004; 15(7–8):891–899.
- Popper R, Kroll RD. Just-About-Right scales in consumer research. *Chem Sense*. 2005; 7(3):3–6.
- Rossi F. Assessing sensory panelist performance using repeatability and reproducibility measures. *Food Quality and Preference*. 2001; 12(5–7):467–479.
- Rothman, L.; Parker, MJ. Just-About-Right (JAR) Scales: Design, Usage, Benefits, and Risks. American Society for Testing & Materials; 2009.
- Shepherd R, Smith K, Farleigh CA. The relationship between intensity, hedonic and relative-to-ideal ratings. *Food Quality and Preference*. 1989; 1(2):75–80.
- Stone, H.; Sidel, JL. *Sensory Evaluation Practices*. San Diego, California: Elsevier Academic Press; 2004.

- van Trijp HCM, Punter PH, Mickartz F, Kruithof L. The quest for the ideal product: Comparing different methods and approaches. *Food Quality and Preference*. 2007; 18(5):729–740.
- Vickers Z, Holton E, Wang J. Effect of yogurt sweetness on sensory specific satiety. *Journal of Sensory Studies*. 1998; 13(4):377–388.
- Vickers Z, Holton E, Wang J. Effect of ideal–relative sweetness on yogurt consumption. *Food Quality and Preference*. 2001; 12(8):521–526.
- Worch T, Dooley L, Meullenet JF, Punter PH. Comparison of PLS dummy variables and Fishbone method to determine optimal product characteristics from ideal profiles. *Food Quality and Preference*. 2010; 21(8):1077–1087.
- Worch T, Le S, Punter P, Pages J. Assessment of the consistency of ideal profiles according to non-ideal data for IPM. *Food Quality and Preference*. 2012a; 24(1):99–110.
- Worch T, Le S, Punter P, Pages J. 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*. 2012b; 26(1):74–80.
- Worch T, Le S, Punter P, Pages J. Ideal Profile Method (IPM): The ins and outs. *Food Quality and Preference*. 2013; 28(1):45–59.
- Worch T, Ennis JM. Investigating the single ideal assumption using Ideal Profile Method. *Food Quality and Preference*. 2013; 29(1):40–47.
- Xiong R, Meullenet JF. A PLS dummy variable approach to assess the impact of jar attributes on liking. *Food Quality and Preference*. 2006; 17(3–4):188–198.

Highlights

- Ideal intensities were not necessarily located at the center of the scale.
- JAR and ideal scaling generated similar estimates of “ Too Little” and “ Too Much”.
- Both scales both identified sweetness and coffee flavor as critical factors in liking.
- The penalty for being too sweet was smaller than being not sweet enough.
- “Too much” and “too little” coffee flavor showed similar influence on liking.

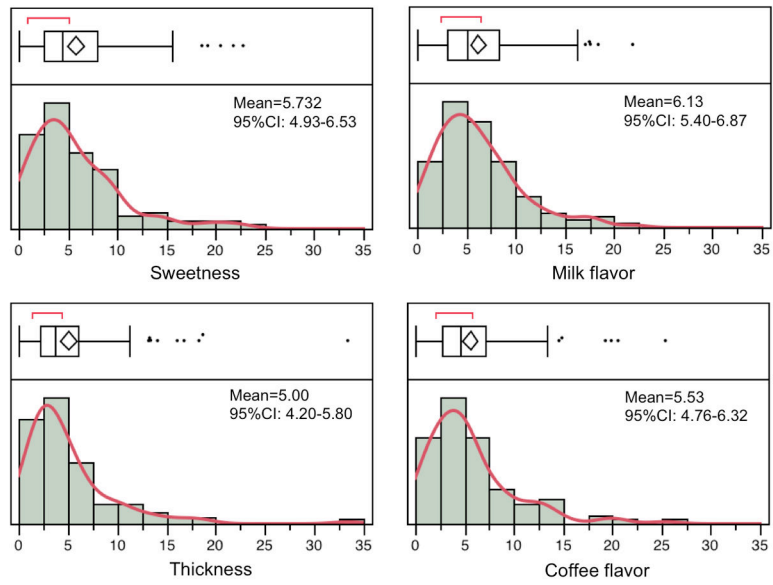


Figure 1. Distribution of standard deviations of individual ideal ratings Data points in the Tukey's box-and-whisker plots beyond the terminus of the whisker (1.5 X IQR) were identified as outliers.

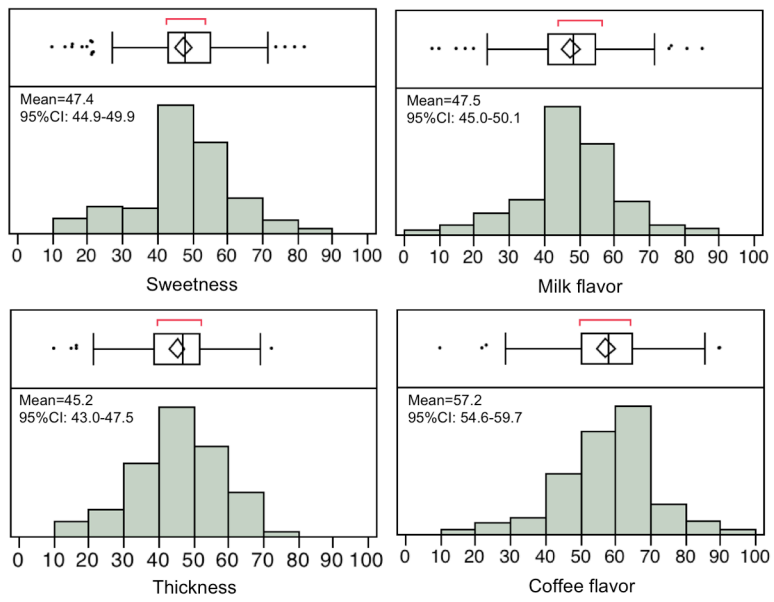


Figure 2. Distributions of ideal intensity ratings using the mean ideal for each participant (n=114)

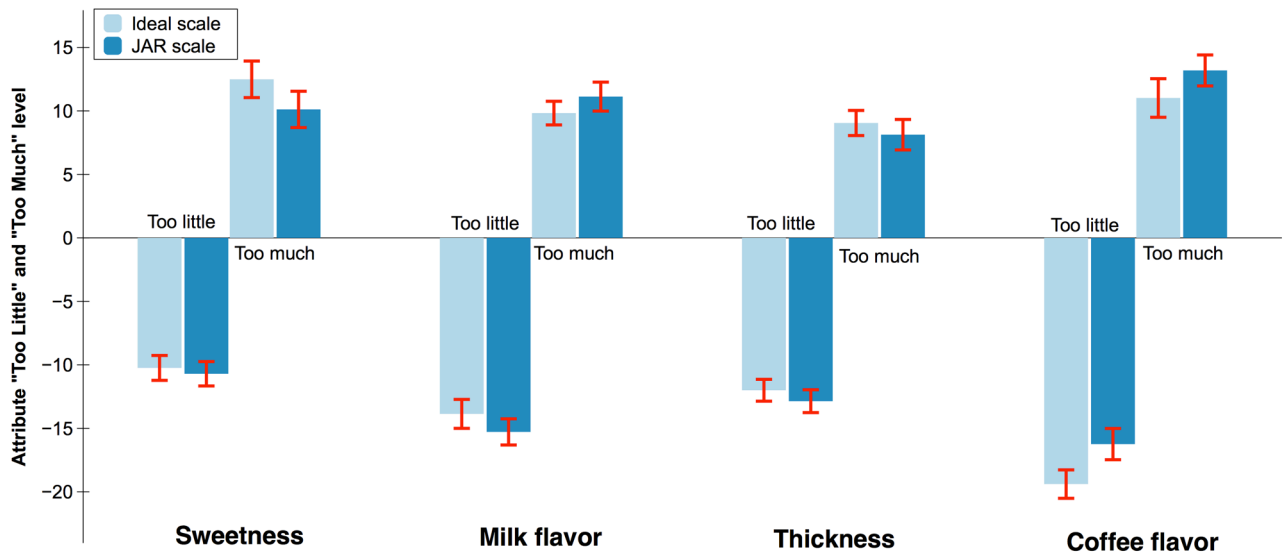


Figure 3. Estimates of “Too Little” and “Too Much” did not differ between the two scaling methods.

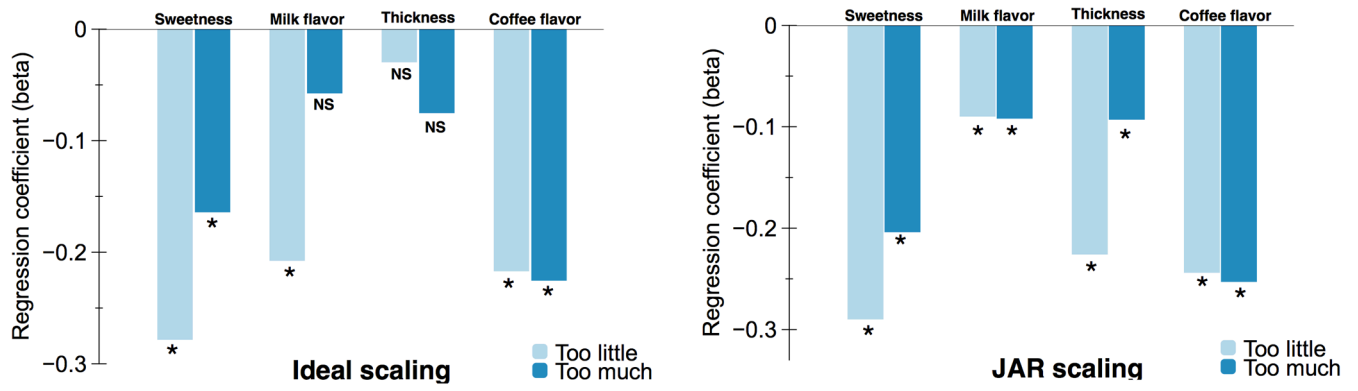


Figure 4.

Influence of attribute “Too Little” and “Too Much” on *liking*. Bars marked with * indicate attribute showed significant influence on liking at $\alpha=0.05$. NS indicates absence of a significant influence on liking.

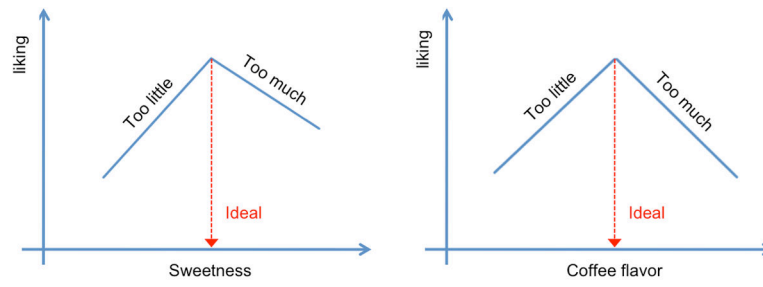


Figure 5.
Asymmetrical impacts of *sweetness* and *coffee flavor* on *liking*.

Table 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 sugar	61.2	58.3

Note: This is a “check all that apply” question. So the sums of percentage for ideal scaling or JAR scaling may exceed 100%.

Table 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.

Table 3

Effect of product and serving order on ideal ratings

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

Table 4

Effects of product and method on “Too Little” and “Too Much”

Performance	Attribute	Term	F-Ratio	P-value	R ² -adj.
Sweetness		Product	9.53 _(19,517.5)	<.0001	
		Method	0.11 _(1,229.9)	0.741	62.4%
		Product*Method	0.73 _(19,517.5)	0.7938	
Milk flavor		Product	2.17 _(19,323.0)	0.0035	
		Method	0.72 _(1,169.9)	0.3974	53.9%
		Product*Method	1.42 _(19,323.0)	0.1139	
Too little		Product	2.63 _(19,461.8)	0.0002	
		Method	3.57 _(1,221.0)	0.06	53.3%
		Product*Method	1.09 _(19,461.8)	0.3605	
Coffee flavor		Product	2.86 _(19,456.8)	<.0001	
		Method	0.48 _(1,193.9)	0.4884	50.4%
		Product*Method	1.20 _(19,456.8)	0.251	
Sweetness		Product	3.72 _(19,136.4)	<.0001	
		Method	1.36 _(1,149.8)	0.2461	79.1%
		Product*Method	1.02 _(19,136.4)	0.4446	
Milk flavor		Product	2.74 _(19,340.2)	0.0002	
		Method	0.79 _(1,160.7)	0.3762	42.6%
		Product*Method	1.03 _(19,340.2)	0.4281	
Too much		Product	3.34 _(19,217.7)	<.0001	
		Method	1.23 _(1,179.5)	0.2695	64.4%
		Product*Method	1.05 _(19,217.7)	0.402	

Performance	Attribute	Term	F-Ratio	P-value	R ² -adj.
		Product	0.99 _(19,185.1)	0.478	
	<i>Thickness</i>	Method	0.35 _(1,140.3)	0.5528	54.1%
		Product*Method	1.48 _(19,185.1)	0.0981	