



Chemo sense

OPINION

Musings on Sensory Crocodiles

Hildegard Heymann

Professor, Sensory Science

Department of Viticulture and Enology

University of California - Davis, USA.

hheyman@ucdavis.edu

In 1979, RoseMarie Pangborn wrote an article with the secondary title 'The Crocodiles are coming.' In this article, written for a short course, she described among other things a quiz that one should subject oneself to prior to doing any sensory testing. She also discussed some aspects of sensory physiology, namely taste as defined by consumer and by the physiologist, issues related to threshold values and (mis)uses of these values, magnitude estimation and unresolved problems with suggested areas for future research in sensory science. I was reminded of this article recently when I had three separate conversations that indicated to me that despite the intervening 25 years - the crocodiles are still coming.

The first conversation was with a professional in industry who had been told that her well-trained and experienced panel was 'useless' because they did not perceive the odor associated with a specific volatile compound in the product. The panel was deemed 'useless' because instrumentally it was known that a specific compound was in the product at a level above threshold. The assumption was that, therefore the panel should be able to perceive it. However, upon further questioning

cont. pg 2

Just-About-Right Scales in Consumer Research

Richard Popper and Daniel R. Kroll
Peryam & Kroll Research Corporation

6323 North Avondale Ave.
Chicago, IL 60631, USA

Richard.Popper@pk-research.com

New product introductions are critical to the growth, continuing success and competitive strategies of packaged goods companies. In order to improve the odds of having a successful launch - whether of an innovative product or a line extension - companies routinely incorporate consumer feedback in the product development process and obtain consumer reactions to product prototypes as they emerge from R&D. Whether it's food, beverage or personal care, the stakes involved are so high that this investment in consumer research not only makes sense, it's probably a necessity, and it can spell the difference between success and failure.

The research tools employed in the pursuit of a winning product formulation run from the simple to the highly sophisticated. The primary objective, however, is the same: provide the product development team with direction on how to increase consumer appeal.

In product categories where sensory properties are important determinants of consumer appeal, one of the most simple and direct ways to solicit feedback from a consumer is to ask whether a product is just right with regard to a certain characteristic or has too much or too little of that

cont. pg 2

INSIDE:

What's Wrong with Sensory?

Wine Sense: New Tool

Program for Paradise

Student Assistance



TM

E-Nose Pty Ltd

Graham Bell and Associates Pty Ltd

Centre for ChemoSensory Research

ISSN 1442-9098

Just-About-Right Scales in Consumer Research continued

characteristic. These "just-about-right" scales (Figure 1) can be effective in research on food and beverages, where consumers, in addition to rating their liking of a product, are asked to evaluate a product on a number of attributes using this question format.

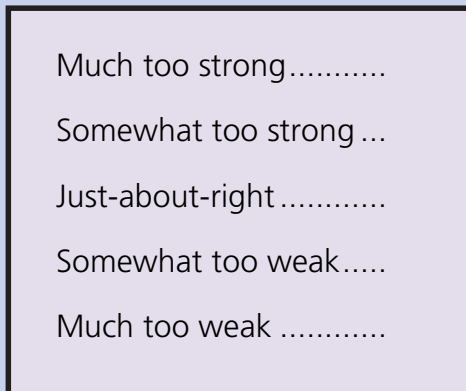


Figure 1. Example of a five-point just-about-right scale used in product evaluation. Other versions of the scale employ three or seven response categories, with the middle category labeled "just-about-right".

For example, in a study of carbonated soft drinks, consumers might be asked to evaluate the prototypes with regard to sweetness, strength of flavor and carbonation level (among other characteristics), indicating each time whether the level is too low, too high or just right. Based on consumers' answers to these questions, the soft drink manufacturer might adjust a prototype's sweetness, flavor and carbonation in an effort to improve its acceptability. Respondents tend to answer just-about-right questions with ease and researchers like the simplicity of the scale. Yet despite their intuitive appeal to researchers and research participants alike, these scales are not without limitations and potential pitfalls, and their results require careful interpretation.

In order to rate sweetness using a just-about-right (JAR) scale, respondents must decide how closely the sweetness of the product they are tasting matches their ideal sweetness. There can often be a disparity between the product

formulations consumers will rate as "just right" and those that they actually like the most. Epler, Chambers, and Kemp (1998) asked to evaluate five lemonades differing in the amount of added sugar. Consumers rated the product using either a JAR scale that ranged from "not sweet enough" to "much too sweet," or rated their liking of the sweetness on a scale that ranged from "dislike extremely" to "like extremely." The optimal sugar concentration was determined in two ways: by identifying the formulation whose average JAR rating was closest to "just right" and by identifying the formulation with the highest average liking score. Using the data reported in that study, Figure 2 shows how the optimal levels can be determined using each measure.

For the JAR scale, the optimal sugar concentration was about 9.5%; for the overall liking scale about 10.5%. While the difference may seem small, it was large enough to make a difference in a preference test. A separate group of

consumers, when presented with the two formulations, preferred the product optimized on the basis of liking over the one optimized on the basis of the JAR scale.

In another study the disparity between optimizing a formulation based on a JAR scale versus a liking scale was even greater. Optimizing the level of aspartame in a fruit drink on the basis of a JAR scale for sweetness predicted an optimal level of aspartame 20% lower than that predicted on the basis of overall liking (Popper, Chaiton & Ennis, 1995).

When products vary in more than just one dimension a similar question arises: is the formulation that maximizes overall liking the same as the formulation for which the sensory characteristics are all just right? In some instances (Moskowitz, Munoz & Gacula, 2003), a product that was just right on all attribute measures was not the same as the product that was liked the most, although it was still an acceptable product; in another instance (Marketo &

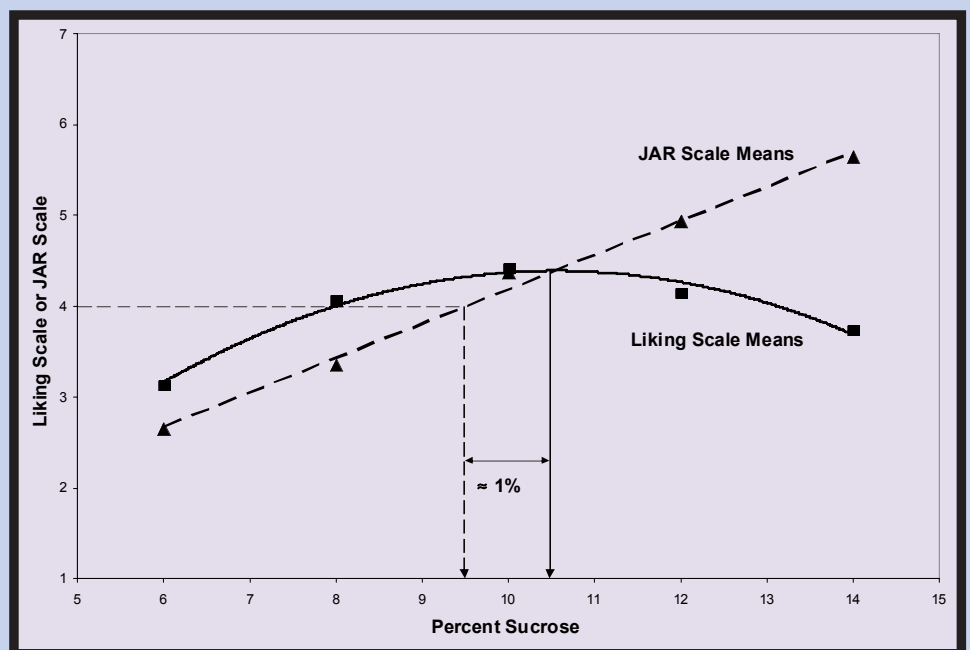


Figure 2. Optimal sucrose level in lemonade determined using a 7-point liking scale (solid line) and a 7-point JAR scale (dashed line). The optimum sucrose level is the level at which the liking curve reaches its maximum or the JAR line equals 4 (just-about-right). The difference in the optimal values is approximately 1%. Source: Epler, et al., 1998.

Just-About-Right Scales in Consumer Research continued

Moskowitz, 2004), the two methods gave similar results.

One possible explanation, advanced by Epler et al., for the discrepancy between JAR and liking scale optima is that JAR scales induce a response bias when the attributes carry certain negative health connotations. In the case of sweetness respondents may perceive a very sweet product as being unhealthful. When tasting such a product, they may say it is "too sweet" because they are aware of the potentially negative consequences of consuming such a product on a regular basis. At the same time, they may actually like the way it tastes, which is reflected in their hedonic ratings. Only one study (Bower & Baxter, 2003) has attempted to confirm this hypothesis by comparing JAR and liking ratings of sweetness among two groups of respondents that differed in their concern with "healthy eating". Unfortunately, the results were inconclusive.

While health concerns may or may not be a source of response bias in the use of JAR scales, experienced researchers know that there are certain attributes that, by their nature, are likely to induce

more, even though their liking ratings may begin to decrease as the level of chocolate chips or the amount of cheese rises above a certain level.

The reverse skew can also occur, say when respondents are asked to rate the bitterness of coffee: the responses will skew towards too much, since bitterness is considered "bad."

However, a certain amount of bitterness may actually be a positive in terms of overall liking, JAR ratings to the contrary.

What direction to product development do JAR ratings provide? Consider the results for a hypothetical product shown in Table 1.

The results suggest that the sweetness and citrus flavor of the product should be increased; less clear is whether the carbonation level should be raised. The results definitely do not tell the product developer how **much** of a change in sweetness, flavor, or carbonation would be required to increase the just right percent. It is tempting to conclude that a bigger increase is needed in the case of citrus flavor than sweetness since the percentage of too low responses is greater for citrus flavor than for

levels of these attributes as just right, leading to a greater percentage of future respondents rating the product too sweet or too high in citrus flavor. Finally, the possible interaction among attributes needs to be considered when making formulation adjustments (increasing the citrus flavor may change the desired level of carbonation).

It is also tempting to conclude from the results that an increase in citrus flavor has the greatest potential to improve the overall acceptability of the product, since this was the shortcoming noticed by the greatest percentage of respondents. But this conclusion could also be erroneous. Consumers might be more tolerant of deviations in flavor level than they are in sweetness, making sweetness the higher priority in terms of reformulation. Furthermore, even though the overall level of satisfaction was greater for carbonation than for the other two attributes, it is possible that for those that considered the carbonation too high or too low, this shortcoming was a bigger detractor than anything else.

Such interpretive difficulties underscore the need for researchers to link the just-about-right ratings to the respondent's level of liking. Using one of several analysis techniques, it is possible to rank order the shortcomings in terms of their importance to overall liking, thereby focusing the attention of product development on the critical attributes. In some cases, an attribute garnering a relatively moderate percentage of complaints (e.g. carbonation too low) may be shown to have a high impact on the overall liking of some respondents.

While a more in-depth analysis can make the results from JAR scales more actionable for the product developer, including JAR scales may still be problematic, as was demonstrated in a study by Popper et al. (2004). In the study, respondents rated their overall liking for four dairy desserts. Some

	Carbonated Soft Drink Ratings		
	Too Low (%)	Just Right (%)	Too high (%)
Sweetness	30	55	15
Citrus flavor	50	40	10
Carbonation	22	65	13

Table 1. Hypothetical ratings of a carbonated soft drink on three just-about-right scales, summarized in terms of the percentage of respondents rating the product just right, too low, or too high.

a response bias. It is hard to imagine that an orange juice could have too much "fresh orange flavor," and many respondents would rate the level of fresh orange flavor as "not enough" regardless of the formulation. Similar biases may exist in the case of the amount of chocolate chips in a chocolate cookie, or the amount of cheese on a pizza. In both instances, researchers can expect respondents using a JAR scale to express a desire for

sweetness. But that is not necessarily the case. The sensitivity of the JAR scale to formulation changes is usually not known and may differ by attribute (Moskowitz, 2004). It might require only a small increase in flavor, but a large increase in sweetness to address the perception that these attributes are too low. And it is not known whether such increases would alienate respondents who currently view the

Just-About-Right Scales in Consumer Research continued

respondents rated only overall liking. Other respondents, in addition to rating their overall liking, rated the products on a series of JAR scales, such as sweetness, thickness and flavor intensity. The study showed that the respondents that answered the JAR scale questions rated their overall liking of the products differently than those that rated only overall liking.

If JAR scales are biasing respondents' overall evaluations of products, then including them in studies designed to measure a product's overall liking may be ill advised. Popper et al. (2004) found that intensity scales, which ask respondents to rate the level of sensory intensity on a scale from low to high, did not have the same biasing effects that the JAR scales did, even though the same attributes were being rated. The difference between the two scale types is that in answering JAR questions respondents need to consider how products differ from an ideal, which may focus them on reasons why they like or dislike a product, something that intensity scales may not. Other research (Wilson & Schooler, 1991) has shown that asking respondents to consider reasons for their preferences may subsequently alter their preference choices.

With the difficulties surrounding the use of JAR scales*, why do they remain so popular? One reason is that alternative research methods can be more costly. Systematically varying a number of key formulation parameters and inferring the optimal formulation from the overall liking responses may require more

prototypes than product development thinks it has the time or money to produce and test. Similarly, formulation direction based on a correlation with intensity ratings, whether collected from consumers or from a trained sensory panel, also requires a fair number of prototypes or in-market products in order to be robust. Compare that approach to one of testing only one or two prototypes (and maybe a competitor) and using just-about-right scales for formulation direction, and the appeal of just-about-right scales for product development is immediately apparent.

JAR scales do not give the specificity of direction that product development often requests, which can lead to inefficient testing-and-retesting in order to get the formulation right. Nevertheless, just-about-right scales, in the hands of knowledgeable researchers and along with the appropriate analyses, can do a just-about-right job of serving as a score card for comparing a number of products and indicating areas where there are major product deficiencies.

* This article discusses some of the limitations and caveats surrounding the use of just-about-right scales. A subcommittee of ASTM Committee E-18 is drafting a detailed guide concerning the benefits and risks associated with the use of just-about-right scales. That document will also include examples of the statistical analyses most appropriate for JAR scales ■

REFERENCES

- Bower, J.A. & Baxter, I.A. (2003). Effects 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.
- Epler, S., Chambers, E & Kemp, K. (1998). Hedonic scales are a better predictor than just-about-right scales of optimal sweetness in lemonade. *Journal of Sensory Studies*, 13, 191-197.
- Marketo, C. & Moskowitz, H. (2004). Sensory optimization and reverse engineering using JAR scales. In Data analysis workshop: getting the most out of just-about-right data, *Food Quality and Preference*, 15, 891-899.
- Moskowitz, H.R. (2004). Just about right (JAR) directionality and the wandering sensory unit. In Data analysis workshop: getting the most out of just-about-right data, *Food Quality and Preference*, 15, 891-899.
- Moskowitz, H.R., Munoz, A.M. & Gacula, M.C. (2003). *Viewpoints and Controversies in Sensory Science and Consumer Product Testing*. Food & Nutrition Press, Trumbull, CT, USA.
- Popper, R., Chaiton, P. & Ennis, D. (1995). Taste test vs. ad-lib consumption based measures of product acceptability. Presented at the Second Pangborn Sensory Science Symposium, University of California, Davis, July 30 - August 3.
- Popper, R., Rosenstock, W., Schraidt, M. & Kroll, B.J. (2004). The effect of attribute questions on overall liking ratings. *Food Quality and Preference*, 15, 853-858.
- Wilson, T.D. & Schooler, J.W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60(2), 181-192.

Receive ChemoSense free

Send "subscribe" message to
g.bell@atp.com.au
(Send "remove" to quit)

