Reprinted from IFPress (2003) 6(3)2,3

Just About Right Scales

Daniel M. Ennis

Background: When Goldilocks considered eating the bears' porridge, she found one "too hot", one "too cold", and one "just right". It is not possible to know from this information whether Goldilocks perceived the porridges to be hot or cold in intensity; we just have relative information about the three porridges. Relative ratings such as just-about-right (JAR), ratings relative to a reference, and ratings of one product relative to another all share a common model. The purpose of this report is to discuss this model and explain how it can be used to interpret just-about-right ratings.

A Model for Relative Ratings: One way of thinking about Goldilocks' decision is that she had some ideal porridge temperature in mind when she tasted the three alternatives. Her first sample was too hot because the temperature exceeded a reference temperature. We do not know if this reference temperature is high or low relative to another person's reference temperature. She may like hot porridges or she may like cold porridges. What we do know is that she is not a machine that produces the same reference temperature every time. Although she generally likes a particular temperature, there is some variation from moment to moment. We assume that her ideal temperature follows a normal distribution of intensities. Similarly, when she tastes a particular porridge, she may perceive different temperatures at different times depending on the particular spoonful that she tastes, the temperature of her mouth and the process by which the sample temperature is transduced to a percept of 'hotness'. We capture these different possibilities by assuming that the porridge hotness values are normally distributed about some mean intensity.

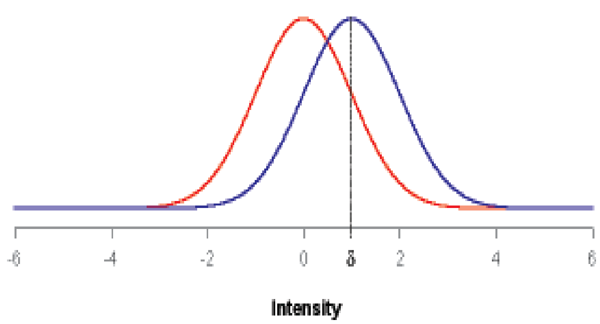

Figure 1: Ideal and product distributions on an intensity scale.

Figure 1 displays two normal distributions, the ideal distribution with zero mean value and a second distribution for the sample with mean value   The ideal mean value is arbitrarily set at zero because we do not know in an absolute sense what her ideal porridge temperature is.

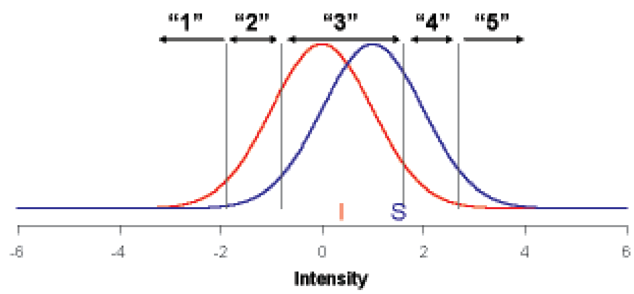

Figure 2a: A rating of 3 is given to sample S.

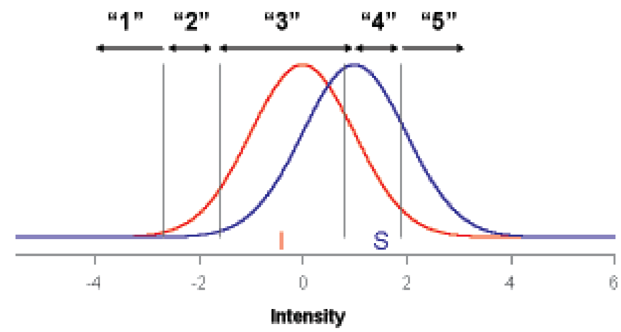

Figure 2b: A rating of 4 is given to sample S.

Goldilocks made her decision to eat Baby bear's porridge based on a single sample of each alternative having found that sample to be closest to her ideal temperature. If she had been asked to rate the porridges, her rating would have reflected this. In order to produce a rating from relative intensity information, boundaries are placed around the momentary ideal intensity, I as shown in Figure 2a, and the rating reported is determined by comparing the momentary sample intensity with the boundaries. Figure 2a shows how ratings from a 5-point JAR scale are generated (1= Much Too Weak; 2 = A Little Too Weak, 3 = Just-About-Right , 4 = A Little Too Strong, 5 = Much Too Strong). A just-about-right rating would be given if the sample intensity, S, fell between the boundaries around the ideal point as it does in Figure 2a. Other ratings are generated based on the other boundary values. The boundary values are symmetrical about the momentary ideal point. The decision boundaries move with the ideal point[1]. Figures 2a and 2b show how a rating of 3 (just-about-right) and 4 (a little too hot) could be given for exactly the same sample hotness but a different momentary ideal.

Table 1. JAR ratings for five products and estimated δ values relative to an ideal point.

Product	Much too Weak	A Little too Weak	Just-about-Right	A Little too Strong	Much too Strong	δ
Product 1	3	10	19	59	9	0.39
Product 2	0	3	43	29	25	1.38
Product 3	0	2	36	30	32	1.67
Product 4	18	26	50	5	1	-1.01
Product 5	1	6	53	25	15	0.87

Scenario: You routinely obtain fragrance intensity JAR data for your air care products from consumer samples. You ou would like to estimate ä values for test products relative to a mean ideal point. You would like to establish a norm for expected values of the JAR scales for an ideal fragrance. This norm could be used to determine whether a particular product in future tests is different from ideal. JAR data are available for five fragrances assessed by 100 consumers each on a 5-point JAR scale. Table 1 displays the data from this experiment.

Model Fitting: A Thurstonian model for JAR data connects the probability of a rating (“1”, “2”,...“*n*”) on an *n*-point scale with the location of product means (δ values) relative to an ideal point at zero along with the location of relative decision boundary values. These values are called relative because they are placed around the momentary ideal values. The counts of the rating values for each product can be used to estimate the parameters of the model using the method of maximum likelihood[2,3]. This method yields the location of product points, decision boundaries, variances of the estimates, and goodness of fit tests. Figure 3 shows the location of the five fragrances relative to an ideal at zero. Fragrances 1, 2, 3, and 5 are stronger than ideal and fragrance 4 is weaker. The decision boundary values are 1.2 and 2.3. The δ values are given in Table 1.

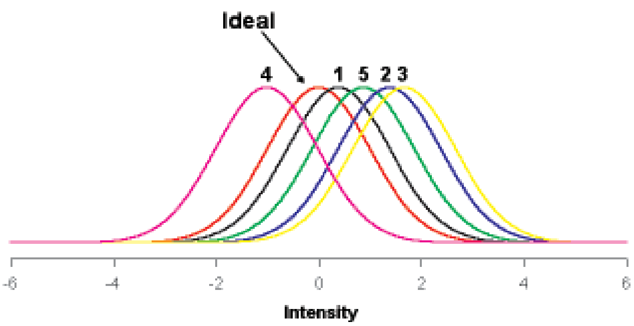

Figure 3: **Scale means of five products relative to an ideal at zero.**

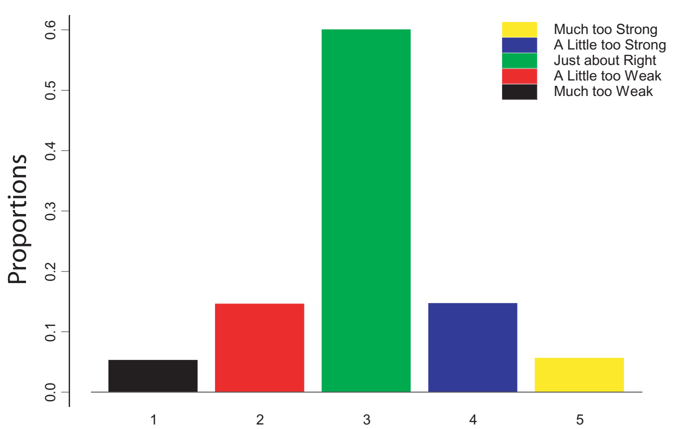

Figure 4: **Proportions of each possible response to an ideal fragrance.**

Estimating an Ideal Norm: From these values we obtain an expected norm of 5% “1’s”, 14.5% “2’s”, 61% "3's", 14.5% "4’s", and 5% "5's" for an ideal fragrance as shown in Figure 4. This norm can be used in future JAR scale testing by comparing product rating frequencies with this norm. For instance, suppose that of 100 judgments for a new fragrance there were 7 “1’s”, 13 “2’s”, 58 “3’s”, 16 “4’s”, and 6 “5’s". Using the norm as expected values, a comparison of the tested fragrance and the ideal is not significant at the 95% level (chi-square with 4 degrees of freedom is 1.46.) Suppose that the rating frequencies for a new fragrance were 10, 20, 40, 20, and 10 for the five categories. Their ratings have a mean of 3 (the just-about-right point) but the frequencies are significantly different from the ideal norm (chi-square with 4 degrees of freedom is 21.4, $p < 0.001$.) An interpretation of this result is that there is both a greater number of consumers who think this fragrance is too strong and also a greater number who think that it is too weak. This information would not be apparent without a norm for the ideal fragrance. An ideal fragrance, as can be seen, has variance that causes the rating to assume values other than 3, the just-about-right point.

Other Relative Scales: Another example of a relative rating method is the relative-to-reference method. In this case products are compared to a reference, such as a consumer’s own brand, to make relative intensity decisions. This type of data can be modeled using the same procedure described for the JAR rating method. The ideal point is replaced by the reference so that the locations of test products and the reference can be obtained. A third type of relative rating method is the rated 2-Alternative Choice (2-AC) method. A special case is the 2-AC method with a ‘no difference’ option. In this case a consumer is instructed to select the strongest of two alternatives on a specific attribute, and then to rate the degree to which the chosen product is strongest. This is also a relative rating method in which one of the products plays the role of reference. One useful outcome of using the model for this type of method is to establish norms for testing differences and preferences with the ‘no difference’ and ‘no preference’ options.

Conclusion: When Goldilocks chose to eat Baby bear’s porridge, she must have perceived its hotness to be closer to her ideal than the alternatives. Without more data from her, we do not know where her ideal point is relative to the three products or what her ideal norm is for a particular rating method. However, with JAR data and an appropriate Thurstonian model, we could estimate these parameters and use them to get the temperature of the porridge “just right.”

References:

1. Ashby, F. G. and Ennis, D. M. (2002). A Thurstone-Coombs model of concurrent ratings with sensory and liking dimensions. *Journal of Sensory Studies,* **17**, 43-59.
2. Dorfman, D.D. and Alf, E., Jr. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals - Rating method data. *Journal of Mathematical Psychology*, **6**, 487-496.
3. Ennis, D.M. (1999). Thurstonian Models for intensity ratings. *IFPress*. **2**(3), 2-3.