Reprinted from IFPress (2022) 25(1) 3,4

Paradoxes in Sensory and Consumer Science

Daniel M. Ennis

Background: An important example of a paradox in the sensory field is *The paradox of discriminatory non-discriminators* discussed by Gridgeman[1] in 1970. A paradox is an apparent contradiction where *apparent* means *appears to be* as opposed to *obvious*. This paradox involved a large observed difference between the choice proportions for the triangular method and the three-alternative forced choice (3-AFC) method involving the same stimuli. Gridgeman called this result a paradox because, unlike a contradiction, he thought that reconciliation was possible. He proposed that the triangular method was a "psychosensorily confusing task." A more insightful resolution was published by Frijters[2] in 1979 and this insight was truly a watershed in the field because it led to an expansion of conventional thinking about the role of perceptual variance in decision making. Frijters explained the result by showing that the two methods agreed extremely well at the level of scaled sensory intensity and that the results from the methods differed because different decision rules were used in the presence of noise. That insight ushered in a whole new vein of research into Thurstonian models[3] that provided a scientific basis for the methodology used in sensory and consumer science. Model predictions could be tested. The now commonly used tetrad method arose out of the implications of that theory.

Equivalence claims are based on a binary choice between two products and involve two bounds set at 45% and 55%. Within these bounds, two products are equivalent. Superiority claims are established when a choice probability (usually preference) exceeds 50%. Unsurpassed claims are made when the choice probability exceeds 45%. An unsurpassed claim combines the concept of equivalence, with a 45% lower bound, and superiority, which has an upper bound of 100%. When an equivalence claim can be supported, there is also support for an unsurpassed claim since equivalence implies that the two products are mutually unsurpassed. The reverse is, however, not true.

The various editions of the *Standard Guide for Sensory Claim Substantiation*[4] (Claims Guide) have always been replete with paradoxes. In a previous technical report[5] we discussed a paradox that an advertiser could claim to be equivalent to a competitor and the competitor could claim superiority over the advertiser with the same data. As with Gridgeman's paradox, this paradox can be explained because different standards are used in the two cases and therefore the two claims are not contradictory. If the Guide were updated to include a consumer-relevant action standard that corresponds to the upper limit for equivalence, then partial or complete elimination of the paradox would occur. It is not necessary, however, to eliminate a paradox, it is just necessary to understand and explain it. In this report, a paradox involving equivalence and unsurpassed claims is explored.

Scenario: You work for a small manufacturer of economy paper towels. Product testing research has shown that the performance of a particular type of your paper towel performs very similarly to a market leader. You believe that a blind preference test among a national sample of loyal users of the market leader's product would find little or no difference between your towels and those of your competitor. Neither type of towel has distinguishing markings that would interfere with blinding. You conduct a national home use test with 320 loyal consumers of your competitor's product of which 158 of them prefer your product and 162 prefer the market leader, closely in line with what you expected. Referring to published tables[6], you find that your data support equivalence since the lower choice count must equal or exceed 158. Since equivalence implies that the two products are mutually unsurpassed, you could also support an unsurpassed claim.

These tables are based on a difference of two binomial functions that exactly corresponds to the hypotheses being tested for equivalence[7]. From this model, you calculate that you have support at the 95% confidence level with a p-value of 0.046 when the null hypothesis of non-equivalence is tested. Checking the recently reissued Claims Guide[4], you find that, according to the Claims Guide, your claim would have required a choice count of 160, rather than 158 to declare equivalence. You calculate the p-value for this outcome and find it to be 0.009, corresponding to the 99% confidence level. Checking the Claims Guide table of critical values for <u>unsurpassed</u> testing, you find the same critical value of 160 is required which means that you could not support an unsurpassed claim either. Surprisingly, all the table values reported for equivalence claims are identical to those for unsurpassed claims and you wonder what model was used to generate these two identical tables.

Hypothesis Testing for Equivalence and Unsurpassed Claims: Assume that X is a random variable representing a discrete measure of the comparative performance of two products and let μ denote its mean. If there is no difference between the products, then μ = 0.5. Equivalence has been defined to mean that μ falls within 0.45 and 0.55. In other words, the null (H_0) and alternative (H_a) hypotheses are:

$$H_0: \mu \le 0.45 \text{ or } \mu \ge 0.55 \text{ and}$$

$$H_a: 0.45 < \mu < 0.55.$$

Corresponding to these hypotheses, equivalence tables are then based on the following published equation for the p-value[7]:

$$p = \sum_{k=0}^{n-m} \binom{n}{k} (0.45)^k (0.55)^{n-k} - \sum_{k=0}^{m-1} \binom{n}{k} (0.45)^k (0.55)^{n-k} \quad (1)$$

where n is the sample size and m is the lower choice count. The critical values of m in the tables are chosen so that p does not exceed 5%.

With n = 320 in an equivalence test, equivalence is established if the advertiser's choice count is 158, 159, 160, 161, or 162 as shown in Figure 1. Counts below or above this set do not reject the null hypothesis of non-equivalence.

Reprinted from IFPress (2022) 25(1) 3,4

For an unsurpassed test, the following are the null and alternative hypotheses:

$$H_0: \mu \le 0.45 \text{ and}$$

$$H_a: \mu > 0.45\,.$$

From these hypotheses, unsurpassed tables are then based on the following equation for the *p*-value:

$$p = \sum_{k=m}^{n} \binom{n}{k} (0.45)^k (0.55)^{n-k} \qquad (2)$$

where *n* is the sample size and *m* is the choice count for the advertiser's product. The values of *m* in the tables are chosen so that *p* does not exceed 5%.

With *n* = 320 in an unsurpassed test, outcomes of at least 160 are required to reject the null hypothesis as shown in Figure 2. The reason for the larger critical value is that in an unsurpassed test there are more possibilities to be considered than in an equivalence test which has two bounds instead of one. This difference leads to the difference in the equations appropriate to their respective null and alternative hypotheses. Since equivalence between two products implies that two products are mutually unsurpassed, this example illustrates a paradox that, like Gridgeman's paradox, has an explanation.

The Paper Towel Scenario: A search through the Claims Guide does not reveal the two models just described. Instead, you find identical Excel and R script applied to the minimum choice count for the equivalence table and the advertiser's choice count for the unsurpassed table. Both refer to equation (2). You wonder why there was not just one table with a note referencing the choice counts and why the correct model for the equivalence case was not disclosed with an explanation for why it was not used. One effect of using the same model for both tests is that you cannot make an equivalence claim in your case using the Claims Guide and you also cannot make an unsurpassed claim. But you should be able to support them both based on the correct equivalence test and also the implication that a supported equivalence claim supports an unsurpassed claim. This brings up again the fact that a paradox is not a contradiction, it is a statement that can be explained. The Claims Guide does not preclude alternative models but requires an explanation if a different conclusion is reached. You decide to conclude from your test that you have correctly supported an equivalence claim, and by implication an unsurpassed claim. Your explanation for obtaining a result different from the Claims Guide is that your analysis uses the correct model for equivalence and the Claims Guide does not.

Conclusion: A paradox is not a contradiction but a statement requiring a deeper understanding than what is obvious. A paradox, such as Gridgeman's paradox, can lead to a cascade of valuable scientific research as its implications are considered. Forcing two different outcomes to agree, as occurs in the current Claims Guide, and obfuscating this fact may discourage valid claims from being made. One such claim could involve a low-cost rival competing with a market leader.

References

1. Gridgeman, N. T. (1970). Re-examination of the two-stage triangle test for perception of sensory differences. *Journal of Food Science* **35**(1), 87-91.
2. Frijters, J. E. R. (1979). The paradox of the discriminatory nondiscriminators resolved. *Chemical Senses* **4**(4), 355–358.
3. Ennis, D. M. (2016). *Thurstonian Models*. Richmond, VA: The Institute for Perception.
4. ASTM International (2021). ASTM E1958-21. *Standard Guide for Sensory Claim Substantiation*. West Conshohocken, PA: ASTM
5. Ennis, D. M. (2020) Claiming superiority and equivalence simultaneously. In Ennis, D. M. and Rousseau, B (Eds.) *Tools and Applications of Sensory and Consumer Science* (pp.74-75). Richmond, VA: The Institute for Perception.
6. Ennis, D. M. (2008). Tables for parity testing. *Journal of Sensory Studies*, **23**(1), 80-91.
7. Ennis, D. M. and Ennis, J. M. (2009). Hypothesis testing for equivalence based on symmetric open intervals. *Communications in Statistics* **38**(11), 1792-1803.

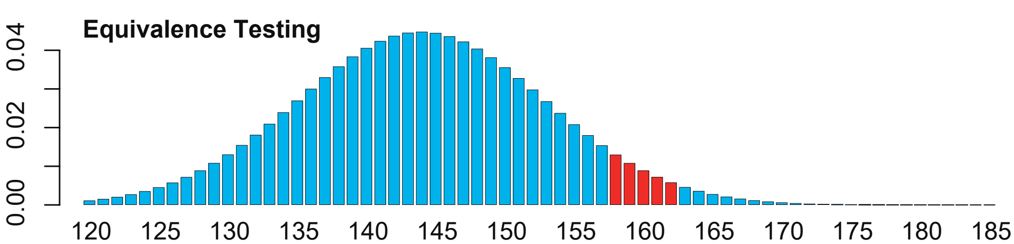

Figure 1. Red bars indicate results supporting equivalence for a binomial distribution with μ = 0.45.

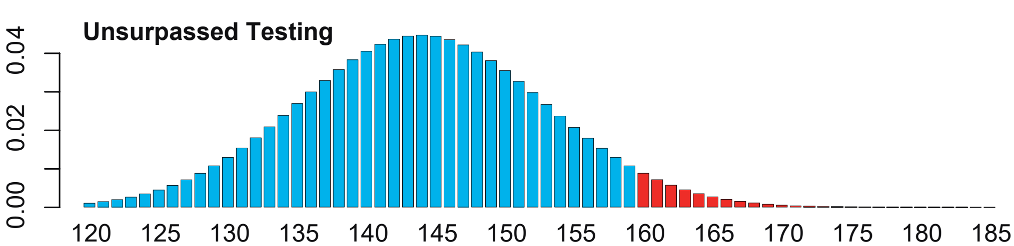

Figure 2. Red bars indicate results supporting unsurpassed for a binomial distribution with μ = 0.45.