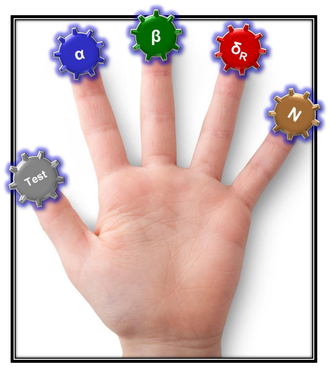


When are two products close enough to be equivalent?

B. Rousseau[1*], M. O'Mahony[2] & R. Ishii[2]

[1]The Institute for Perception, Richmond, USA, [2]University of California, Davis, USA

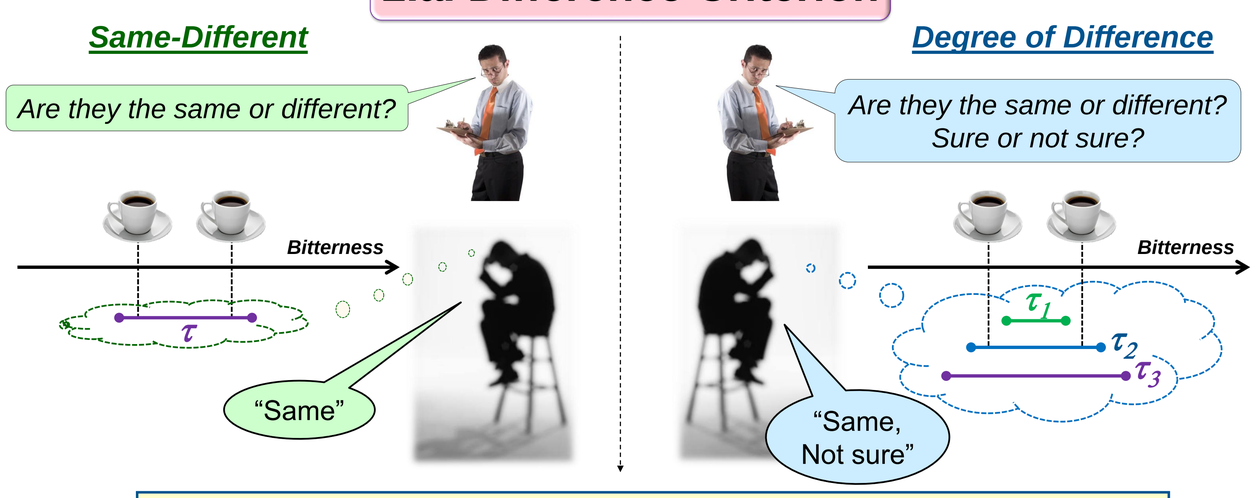
1. Introduction

- In quality control, a product and its 'match' will always be different
- Before making a manufacturing or formulation change, the question is whether they are 'close enough'
- An answer requires the measure of consumer relevance to establish a threshold (δ_R)
- Once the threshold has been established, suitable sample size estimations can be made to maximize confidence in research results
  - e.g.: $\alpha = 5\%$, Power = $1 - \beta = 80\%$, $\delta_R = 1$, { Triangle → N = 220; Tetrad → N = 65 }
- This investigation compared the same-different test and the paired preference test in terms of their predictions for consumer relevance (δ_R)

2. Background

- Thurstonian modeling can quantify the size of the sensory difference (δ or its experimental estimate d')
- Using this standardized measure of difference, δ_R can be investigated

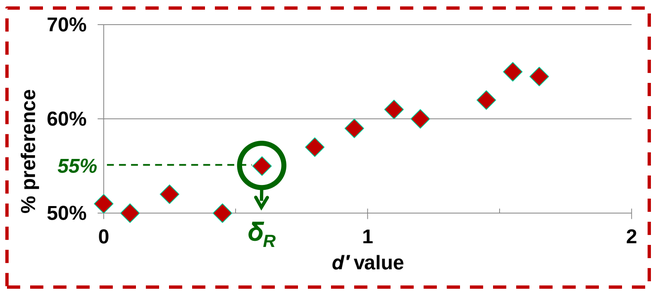
2.a. Difference Criterion



- Analysis estimates
  - d', size of the difference between the products
  - τ, size of the psychological criterion used to generate an answer
  - τ has the same unit as d' and can be used as a measure of δ_R ('close enough')

2.b. Preference Testing

- A relationship between d' and preference is built
- A threshold for a preference of relevant size is set, e.g., 55%
- The corresponding d' value can then be used to set δ_R



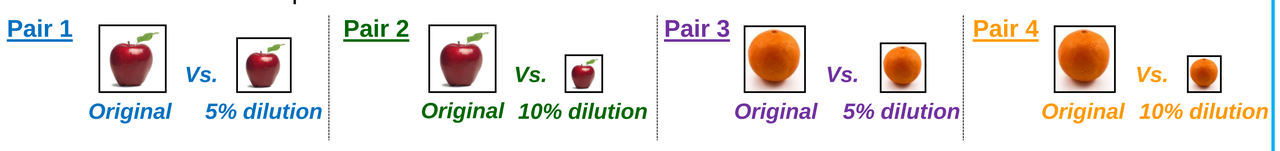
3. Research Objectives

- Conduct degree of difference tests between various pairs of stimuli and investigate the stability of the τ criteria across pairs
- Compare the δ_R value predictions based on degree of difference and preference testing
- Make a recommendation on which approach might be more suitable to establish a consumer relevance threshold

4. Material & Methods

4.a. Subjects & Stimuli

- Subjects: N = 256 (126 males, 130 females, average age 24.8 years old)
- Stimuli: Cascadian Farms organic fruit juices (generously donated by General Mills)
  - Apple
  - Orange
- Four stimulus pairs



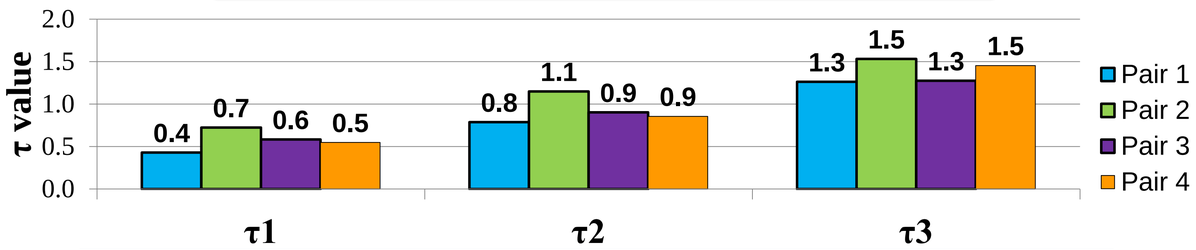
4.b. Procedure

- **Each consumer performed**
  - One *apple juice* comparison (Pair 1 or Pair 2): one paired preference and two degree of difference (one identical pair, one different pair)
  - One *orange juice* comparison (Pair 3 or Pair 4): one paired preference and two degree of difference (one identical pair, one different pair)
  - Number of evaluations per pair and protocol: 128
- **Instructions**
  - Paired preference: "Which sample do you prefer or do you have no preference?"
  - Same-different: "Do you think the samples are the same or different? Are you sure or not sure?"

Preference | DOD

5. Results

5.a. Stability of τ criteria across pairs



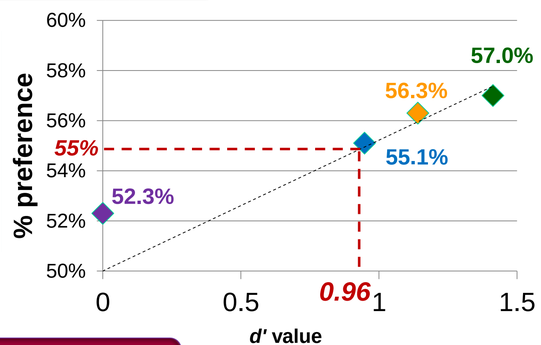
- For each τ level, the τ values are not significantly different ($p > 0.05$)
- The criterion sizes were stable across the different pairs compared with the DOD

5.b. DOD δ_R

- The limit between "Same not sure" and "Different not sure" (τ_2) can be used for δ_R
- Taking the average of the four values, we find that $\delta_R = 0.92$

5.c. Preference test δ_R

- The d' values are those calculated from the DOD
- Plotting their relationship against the preference results uncovers the relationship
- Setting the threshold at 55% preference (a value often used in the industry), the corresponding δ_R is estimated at 0.96



6. Conclusions

- τ criteria are stable across pairs of stimuli and provide a measure of consumer relevance
- Interestingly, both DOD and preference tests provided the same estimate of δ_R
- This indicates that consumers will start exhibiting a preference when they start perceiving the products to be different
- Sample size calculation: $\alpha = 5\%$, Power = $1 - \beta = 80\%$, $\delta_R = 0.96$, { 2-AFC → N = 28; Triangle → N = 251; Tetrad → N = 78 }
- The preference approach is preferred due to the greater statistical power of the procedure
- Further research is needed to investigate this relationship with other types of stimuli

7. References

- Rousseau, B. (2015). Sensory discrimination testing and consumer relevance. *FQP*, **43**, 122-125.
- Ennis, J., Rousseau, B., and Ennis, D. (2014). Sensory difference tests as measurement instruments: A review of recent advances. *JSS*, **29**, 89-102.
- Ishii, R., O'Mahony, M., & Rousseau, B. (2014). Triangle and tetrad protocols: Small sensory differences, resampling and consumer relevance. *FQB*, **31**, 49-55.
- Ennis, J. M. & Jesionka, V. (2011). The power of sensory discrimination methods revisited. *JSS*, **26**, 371-382.
- Ennis, D. M. (1993). The power of sensory discrimination methods. *JSS*, **8**, 353-370.