Image credit: wikipedia.org

- *“ Equipped with our five senses, we explore the universe ... and call the adventure science”*

  - Edwin Hubble

Society of Sensory Professionals

Jersey City, New Jersey

October 11th, 2012

Technical Workshop – An Exploration of Tetrad Testing

John Ennis

The Institute for Perception, Richmond, VA, USA

John Cowden

General Mills, Minneapolis, MN, USA

Karen Garcia

Symrise, Teterboro, NJ, USA

Pieter Punter

OP&P Product Research, Utrecht, The Netherlands

Discrimination Testing

- Discrimination testing as important as ever:
  - Compliance with health initiatives
  - Cost reductions
  - Changes to ingredients, processes, packaging, handling, etc.
  - Quality control
- Three challenges:
  1. Identify sensitive methods for unspecified testing
  2. Measurement:
     a) Quantify sensory differences
     b) Understand precision in measurement
  3. Determine size of meaningful difference

The Tetrad Test - Methodology

- Four samples presented:

“Group the stimuli into two groups of two samples based on similarity”

- Six possible presentation orders: **AABB, ABAB, ABBA BBAA, BABA, BAAB**
- Guessing probability = 1/3

The Tetrad Test - History

- Mentioned by Lockhart (1951) and Gridgeman (1954)
- Revisited by O'Mahony, Masuoka, & Ishii (1994)
- First experiments:
  - Masuoka, Hatjopolous, & O'Mahony (1995)
  - Delwiche & O'Mahony (1996)
- First theoretical analysis:
  - Ennis et al. (1998)
- Support for Tetrad testing in IFPrograms™ (2009)
- Sample size tables published by Ennis & Jesionka (2011)
- Detailed comparison with Triangle test by Ennis (2012)
- Large-scale comparison with Triangle test by Garcia, Ennis, & Prinyawiwatkul (2012)

Overview of Talks 1 & 2

- **Talk 1: An Industry-Based Comparison of the Triangle and Tetrad Tests to Managing Product Reformulation Risk**
  - John Cowden, *General Mills*, Minneapolis, Minnesota, USA
  - Amalie Kurzer, *Brigham Young University*, Provo, Utah, USA
  - Norton Holschuh, *General Mills*, Minneapolis, Minnesota, USA
  - Suzanne Pecore, *General Mills*, Minneapolis, Minnesota, USA

- **Talk 2: Tetrad Testing in Consumer Research**
  - Pieter Punter, *OP&P Product Research*, Utrecht, The Netherlands
  - Lidewij Verweij, *OP&P Product Research*, Utrecht, The Netherlands
  - Bert Borggreve, *H.J. Heinz B.V.*, Zeist, The Netherlands

Overview of Talks 3 & 4

- **Talk 3: A Large-Scale Experimental Comparison of the Tetrad and Triangle Tests in Children**
  - Karen Garcia, Symrise, Teterboro, NJ, USA
  - John Ennis, *The Institute for Perception*, Richmond, VA, USA
  - Daniel Ennis, *The Institute for Perception*, Richmond, VA, USA
  - Witoon Prinyawiwatkul, *Louisiana State University*, Baton Rouge, LA, USA

- **Talk 4: The Sensitivity of the Tetrad, Triangle, and Degree of Difference Tests**
  - John Ennis, *The Institute for Perception*, Richmond, VA, USA
  - Rune Christensen, *Technical University of Denmark*, Lyngby, Denmark
  - Benoît Rousseau, *The Institute for Perception*, Davis, CA, USA
  - Daniel Ennis, *The Institute for Perception*, Richmond, VA, USA

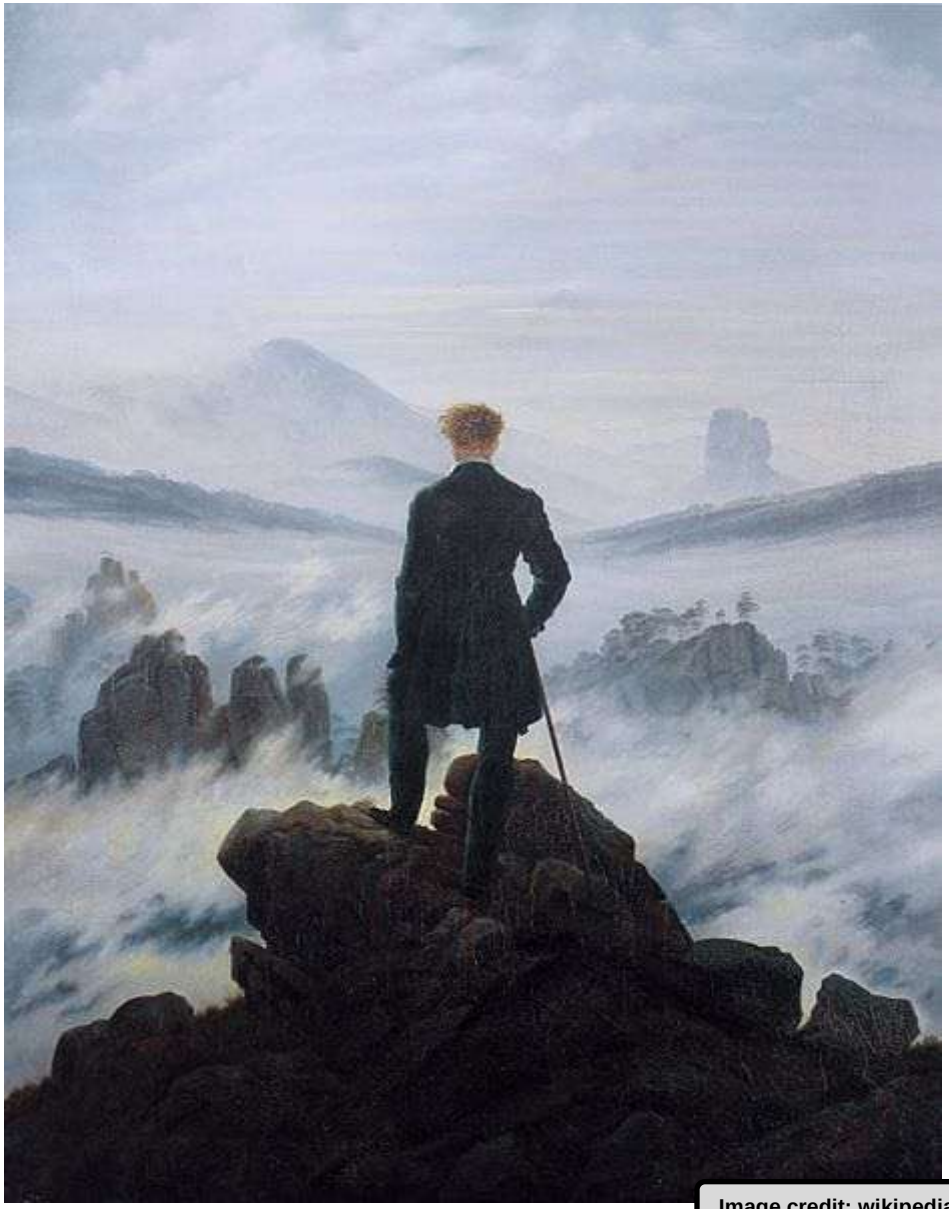


Image credit: wikipedia.org

John Cowden

Innovation Project Manager at General Mills

- M.S. in Food Science from Oregon State University
- As Senior Sensory Scientist, provided sensory and shelf-life consultation to R&D, Quality, Consumer Insights and Marketing
- Led Product Guidance and Insights for the Cereal Platform
- Helps teams identify and bring to life new opportunities

Tetrad vs. Triangle
An Industry Perspective

John Cowden, Suzanne Pecore, Nort Holschuh, Amalie Kurzer

October 2012

Society of Sensory Professionals, Jersey City

John.cowden@genmills.com

Talk Overview

- Background/test primer
- Sensitivity comparison
- Peek into reproducibility
- Tetrad to manage risk

PGI
Product
Guidance &
Insights

Background

- General Mills has a long standing history of using discrimination testing to guide product and processing reformulation changes.
- To date, triangle testing has been used to manage risk for blind product changes.
- Though the triangle method is inexpensive, obtaining adequate product and maintaining a large pool of motivated panelists is challenging.
- Tetrad shows promise to replace triangle methods and overcome current challenges of triangle.

Test Primer, Triangle vs. Tetrad

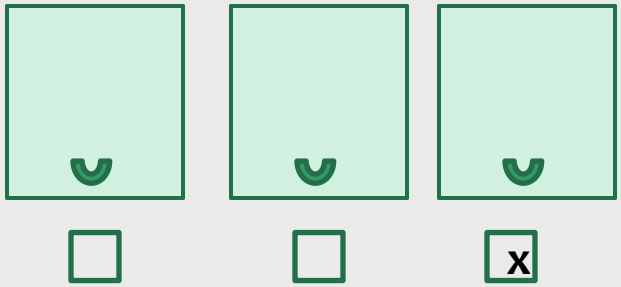
Product Guidance & Insights

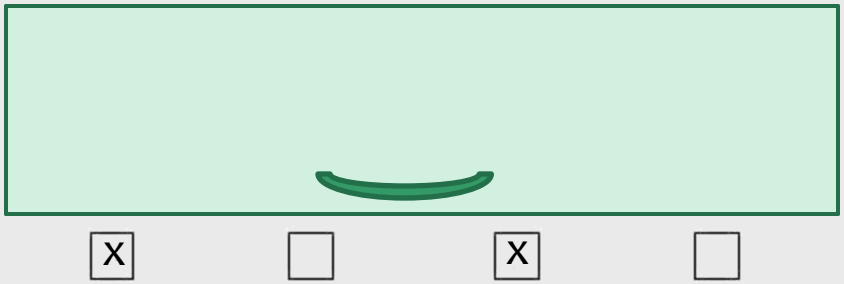
Triangle

- Panelists select the different sample
- **Sequential monadic**
- Outcome is who got pairing correct, p=1/3 by luck

Tetrad

- Panelists group 4 samples into 2 pairs
- **Side-by-side comparison**
- Outcome is who got pairing correct, p=1/3 by luck



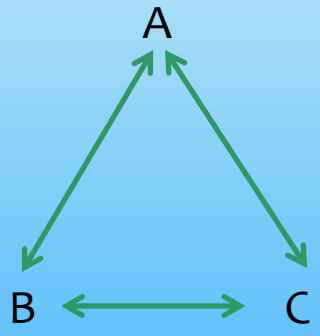


Comparing The Psychological Task

Product Guidance & Insights

- Triangle

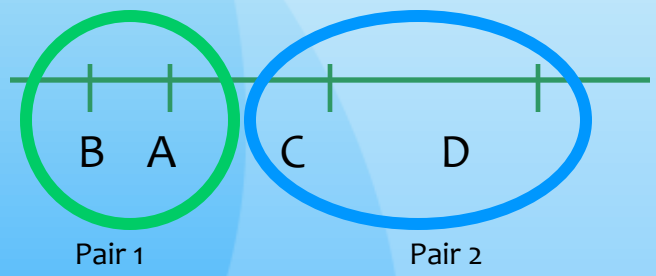
Relies on comparisons



VS

- Tetrad

Uses an easier linear approach

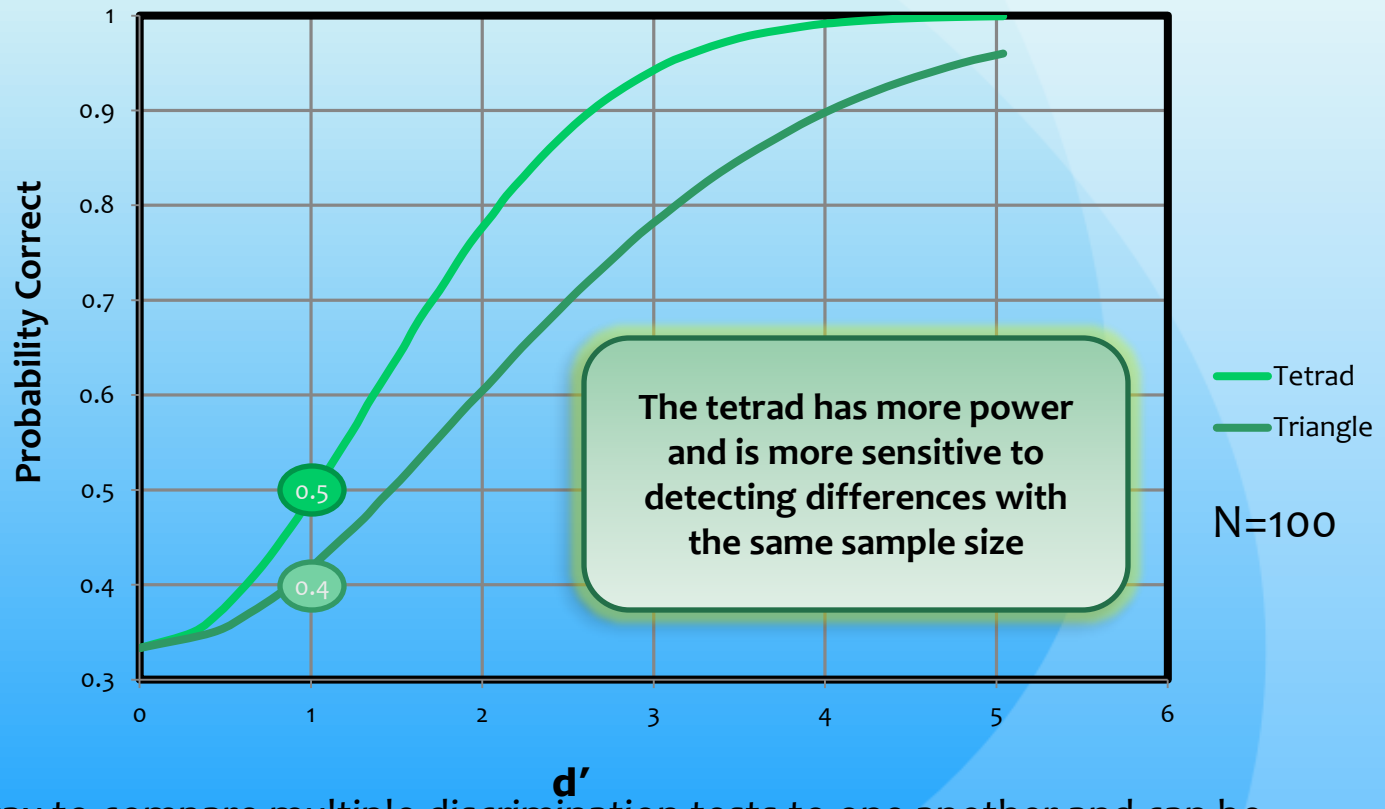


Which pair is most alike?

What is the order?

Respondents More Likely To Find A Difference In Tetrad

Product Guidance & Insights



d' is the way to compare multiple discrimination tests to one another and can be thought of as the amount of difference between test and control products

David HA, Trivedi MC. Blacksburg, Va.: Virginia Polytechnic Insti; 1962. Pair, triangle and duo–trio tests. Technical report nr 55, Dept. of Statistics

Ennis, J. M., Ennis, D. M., Yip, D. and O'Mahony, M. (1998). Thurstonian models for variants of the method of tetrads. *British Journal of Mathematical and Statistical Psychology*, **51**(2), 205-215.

Comparing Sensitivity in Practice

Products Run As Both A Triangle (Δ) And A Tetrad (□)

Product Guidance & Insights

Product	Difference
Cereal 1	Flavor
Cereal 2	Texture/Flavor
Cereal 3	Texture
Baked Good 1	Texture
Baked Good 2	Flavor
Dairy 1	Flavor
Dairy 2	Flavor
Spicy Meal 1	Flavor
Spicy Meal 2	Flavor

Tetrad Is Consistently More Sensitive Than Triangle

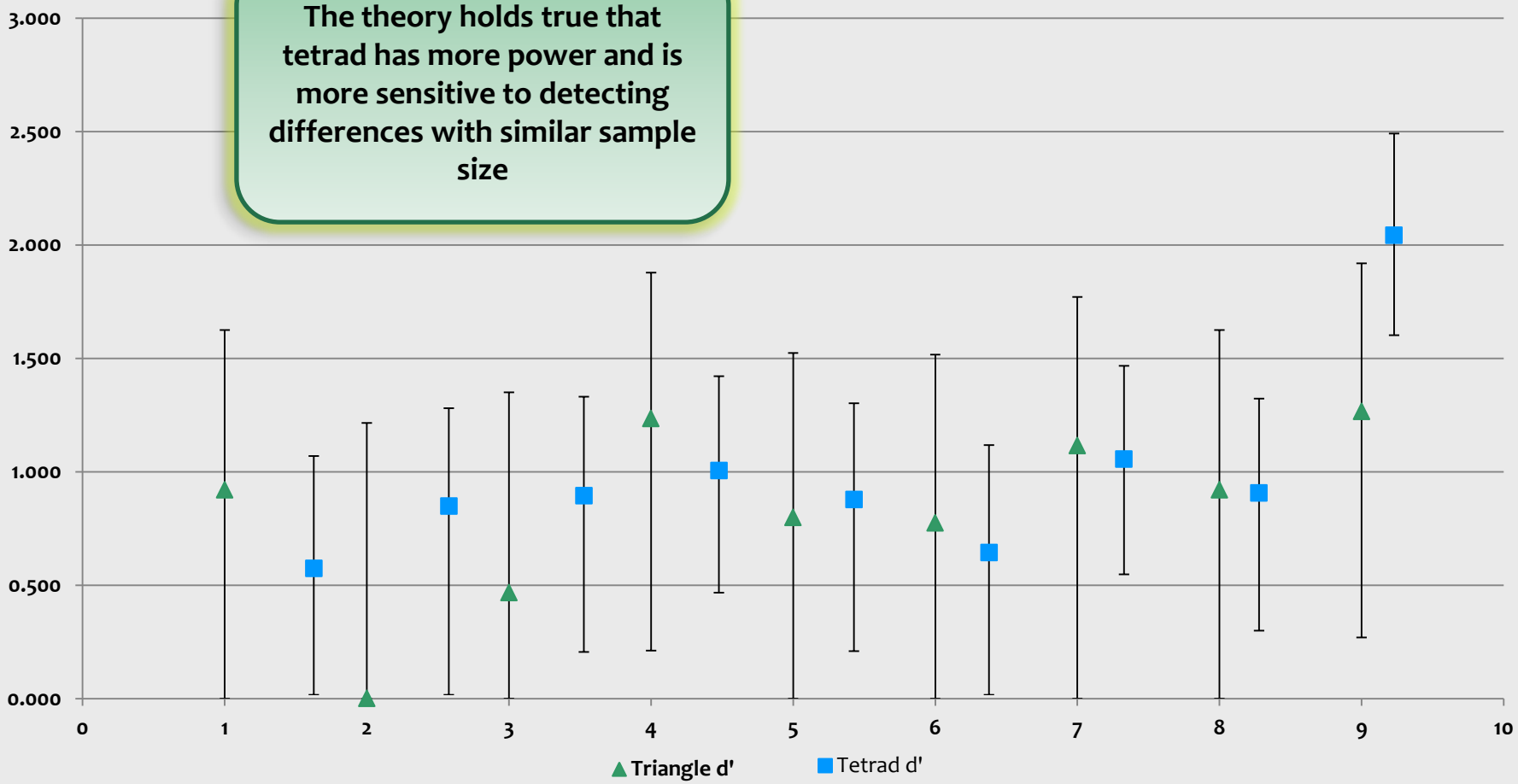
Product	Method	True Discriminators	Sample Size
Cereal 1	Triangle Δ	11%	69
	Tetrad □	8%	72
Cereal 2	Triangle Δ	0%	67
	Tetrad □	18%	72
Cereal 3	Triangle Δ	3%	68
	Tetrad □	19%	67
Baked Good 1	Triangle Δ	19%	70
	Tetrad □	24%	69
Baked Good 2	Triangle Δ	8%	72
	Tetrad □	19%	72
Dairy 1	Triangle Δ	8%	70
	Tetrad □	10%	72
Dairy 2	Triangle Δ	8%	70
	Tetrad □	26%	72
Spicy Meal 1	Triangle Δ	11%	69
	Tetrad □	20%	63
Spicy Meal 2	Triangle Δ	19%	67
	Tetrad □	68%	66

PGI
Product Guidance & Insights

Tetrad No More Fatiguing Than Triangle

Product Guidance & Insights

The theory holds true that tetrad has more power and is more sensitive to detecting differences with similar sample size



Tetrad Is Repeatable And Conservative

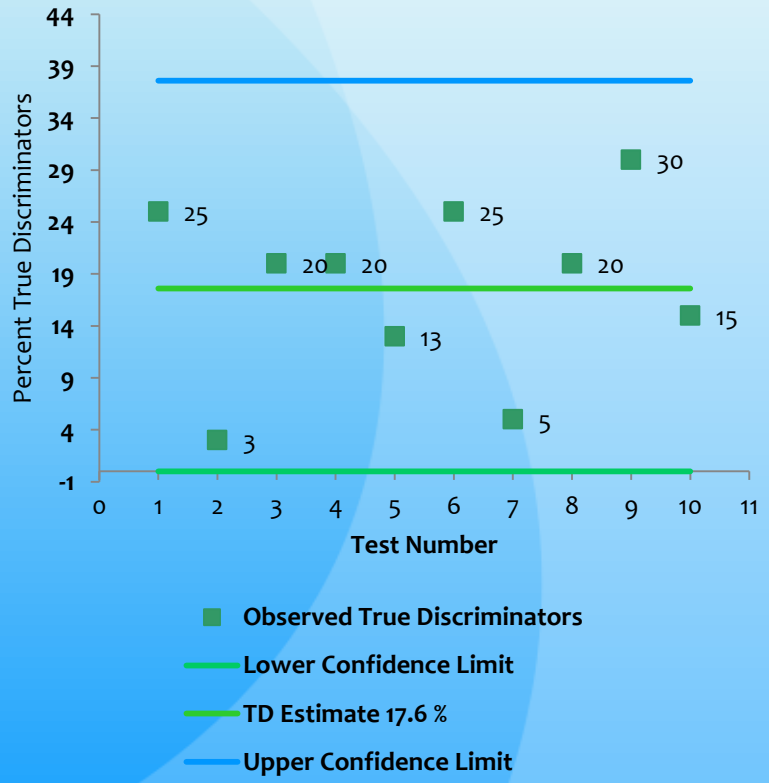
Product Guidance & Insights

Repeated Tetrad

Rep	N	Correct	True Discriminators
1	30	15	25
2	31	11	3
3	30	14	20
4	30	14	20
5	31	13	13
6	30	15	25
7	30	11	5
8	30	14	20
9	30	16	30
10	30	13	15
Average			17.6

Chi-square test for differences in true discriminator across tests is not significant (p =0.9314)

Repeated Tetrad

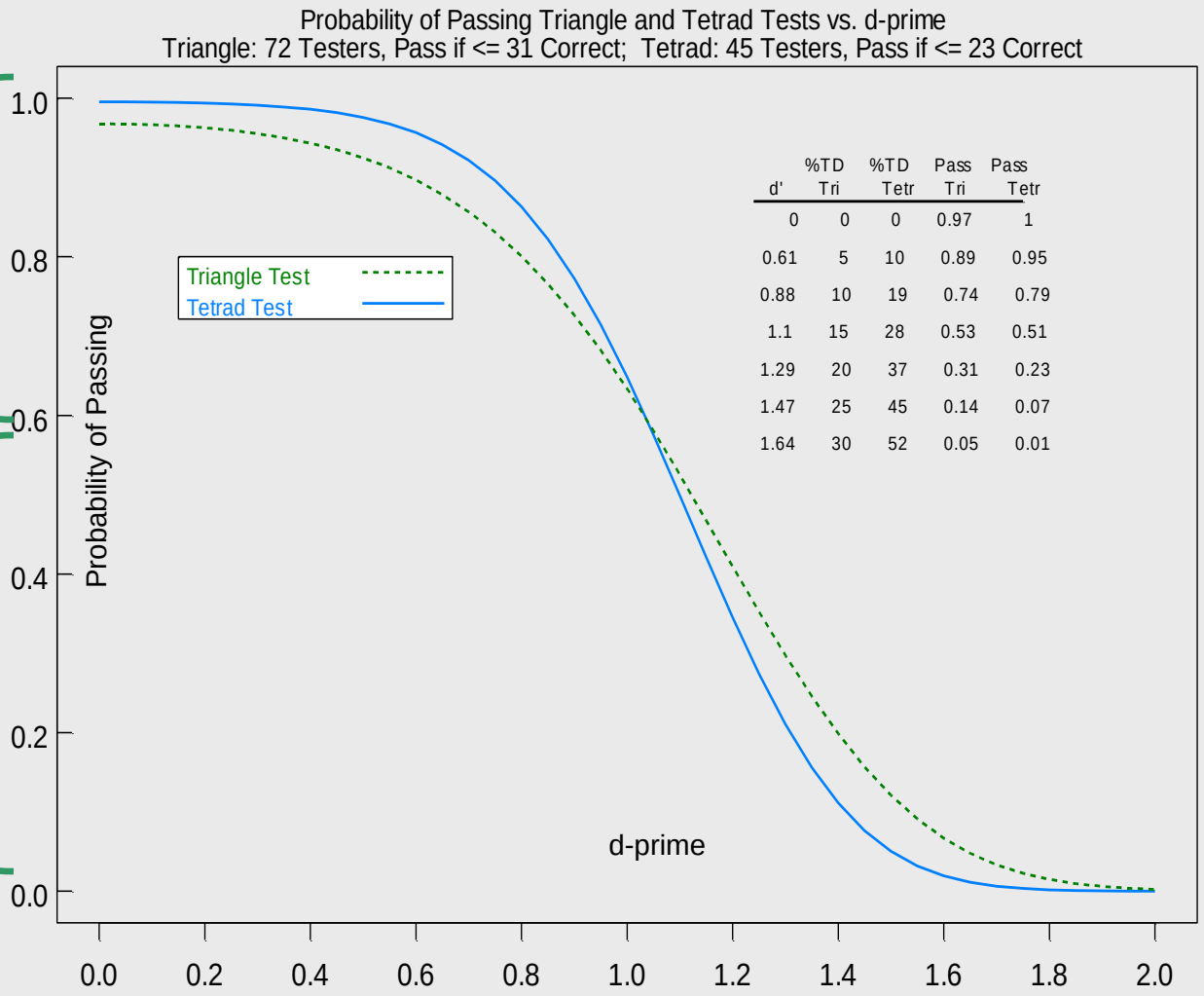


Tetrad Manages Risk Better Than Triangle With Fewer Respondents

Product Guidance & Insights

Tetrad less likely to send R&D back to reformulate when they don't need to (type I error)

Tetrad less likely to pass a change when it should have failed (type II error)



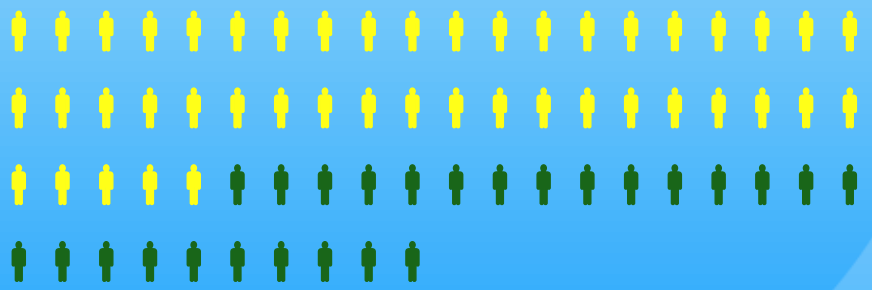
d'	%TD Tri	%TD Tetr	Pass Tri	Pass Tetr
0	0	0	0.97	1
0.61	5	10	0.89	0.95
0.88	10	19	0.74	0.79
1.1	15	28	0.53	0.51
1.29	20	37	0.31	0.23
1.47	25	45	0.14	0.07
1.64	30	52	0.05	0.01

Implications for Replacing Triangle

Product Guidance & Insights

- **The tetrad requires fewer panelists for the same risk profile as current triangle.**
  - Less product required for test = easier for R&D to make samples
  - Fewer respondents = less employee panelist and testing time
  - Less complex for lab to execute = increased testing capacity

Benefit of Tetrad Fewer Respondents



Tetrad n=45

Triangle n=72

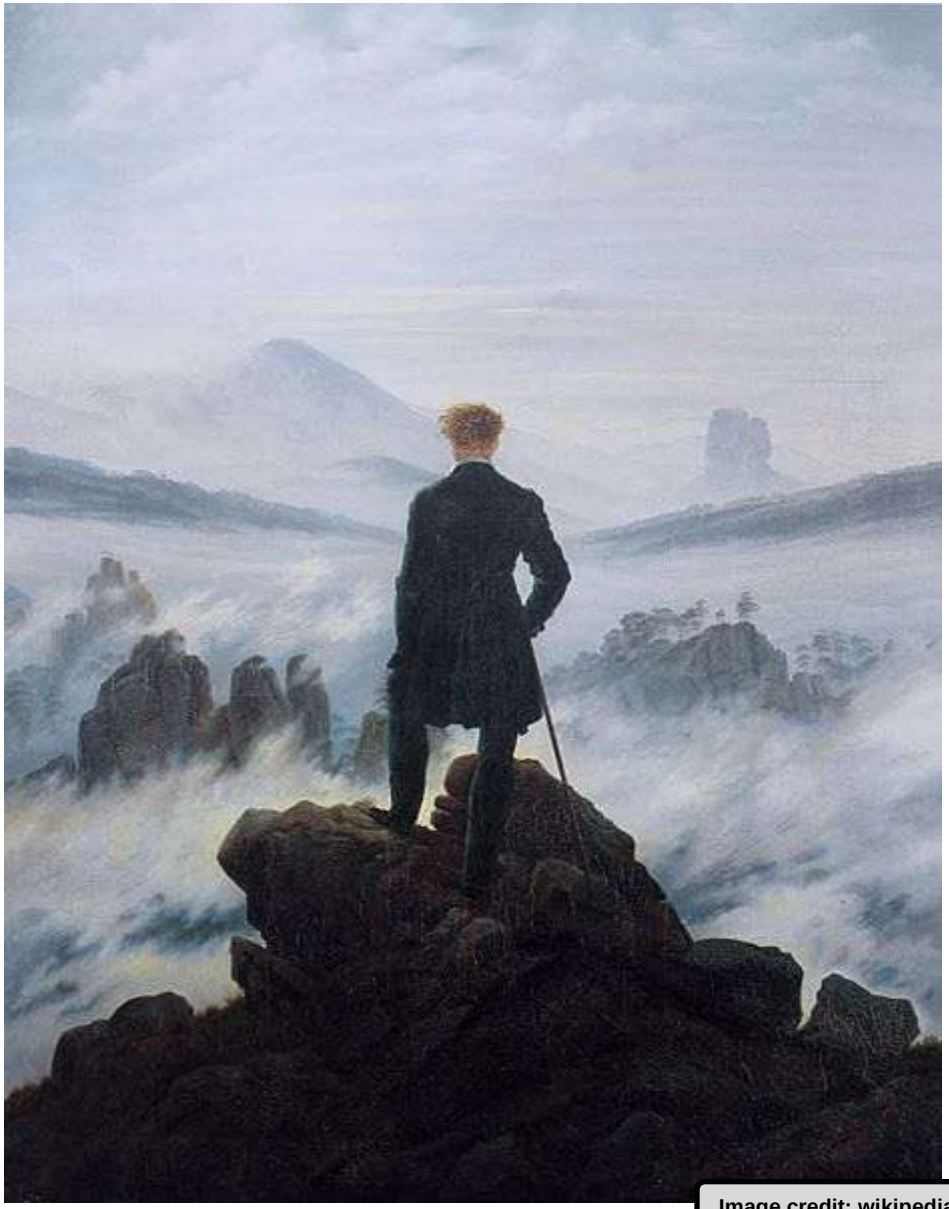


Image credit: wikipedia.org

Pieter Punter

Research Director of OP&P Product Research

- Studied Sensory Psychology in Utrecht, The Netherlands
- Worked for 10 years at the University of Utrecht doing fundamental and applied olfactory research
- Developed the Ideal Profile Method to guide product development
- Consults for a wide variety of food companies in Europe
- Co-founder of the Sensometrics society
- Active member of the Dutch, European, and American Sensory groups

Tetrad testing in Consumer Research

defining benchmarks

Pieter Punter, Lidewij Verweij
OP&P Product Research
Utrecht, The Netherlands

Bert Borggreve,
H.J.Heinz, Zeist, The Netherlands

The unspecified Tetrad and naïve consumers

Two main questions:

- is the method feasible
- how sensitive are consumers?

  - can they detect differences between different products?
  - can they detect differences between formulations from different suppliers?
  - can they detect differences between changes in the formulations?
  - can we define benchmarks?

Benchmarks for d' (delta)

- delta can vary between 0 (identical products) and 3+ (very different products)
  - below value X we assume that their is hardly any perceptible difference
  - above value Y we regard the two samples as clearly different

- somewhere between these two values we would like to draw the line between 'same' and 'different'

- defining benchmarks
  - what is delta for 'dummies' (identical products)?
  - what is delta for really different products (different brands, different formulations)
  - what is delta in case of recipe changes, supplier changes etc?

Experimental

- minimal requirements:
  - sample size and power
    - » the sample size for 85% power and delta=1 is 75 *
  - users of the product category in question
  - all are naive consumers recruited from the OP&P database
  - they will assess 3 or 4 tetrads in 45 or 60 minutes

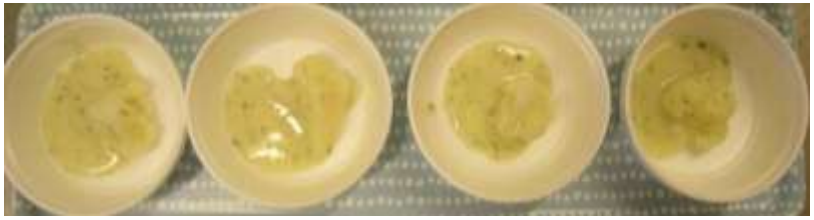
- in the past months, more than 120 tetrads have been executed with 60-85 consumers each

*** for a similar power, more than 200 subjects would be needed with the triangle test**

Tetrads with soups, sauces, pastas and meals

- soups

- sauces



- pastas& meals

Performing 123 tetrads

Question 1: what is delta for identical products (differing in complexity in terms of product and preparation)

- dummy tests with
  - » **same soups**
  - » **same mixes**
  - » **same ketchup**
  - » **same pasta's**

Question 2: what is delta for different products

- tests with products from different brands
  - » **Twix/Lidl (candybar)**
  - » **A-Brand/Private label (instant soups)**
  - » **A-brands/Private labels (mixes for pasta or beans)**

Question 3: what is delta for a supplier switch or a recipe change?

- » **supplier switch for soups, mixes, pasta's**
- » **recipe change for soups, mixes, pasta's**

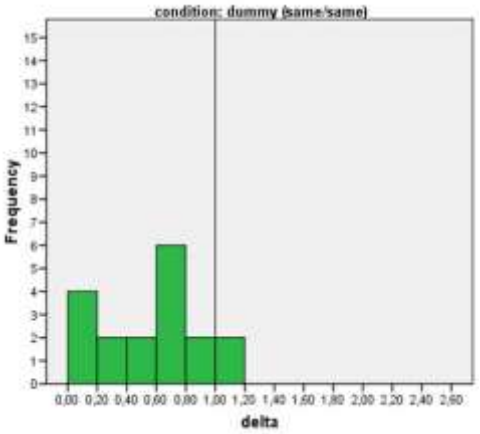
The results

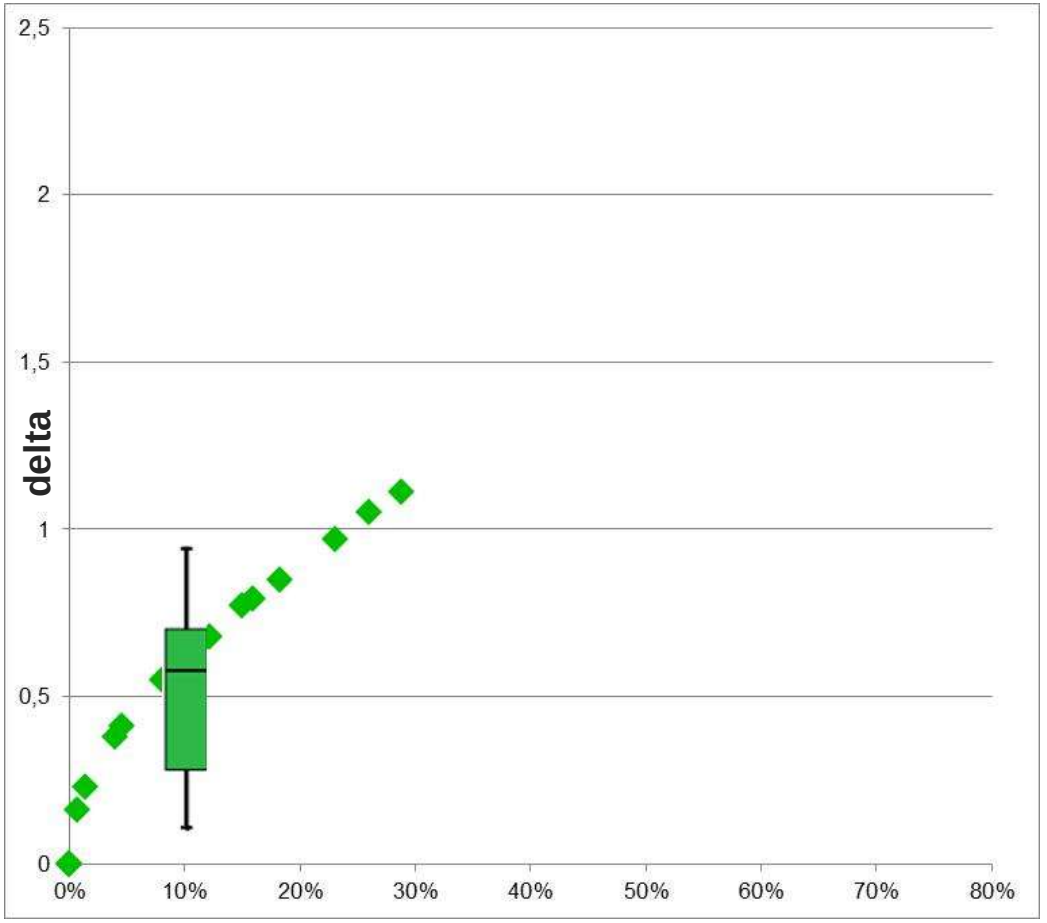
Delta's for dummies (n=18)

how similar are the same products?

- the median value of delta is 0,61
- the delta's range between 0 and 1,2

- pasta meals and soups from the same brand with fresh ingredients are not always identical due to difficulties in preparation, small differences in cooking time and differences in ingredients
- making one batch and splitting is easier than making two different batches with the same ingredients



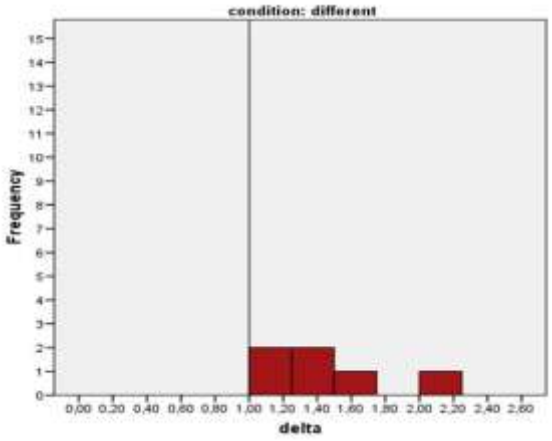


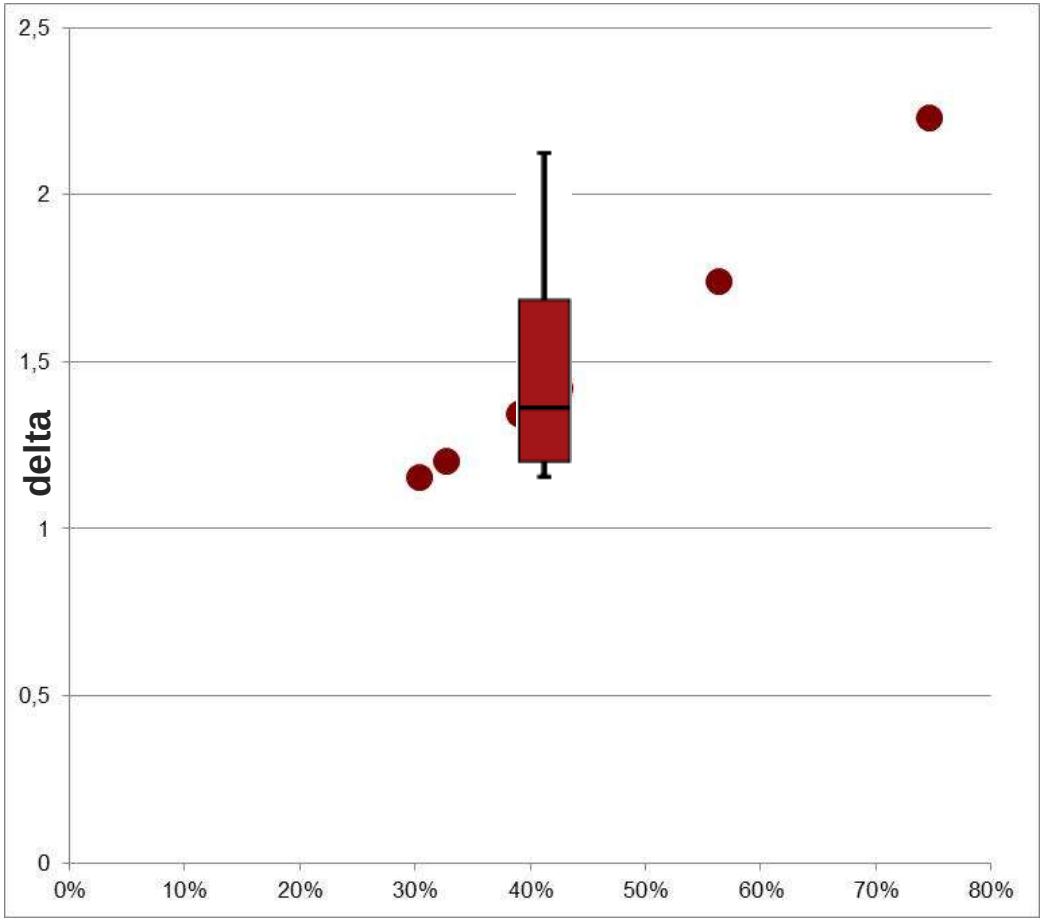
Delta's for different products (n=6)

how different are different products?

- the median value of delta is 1,38
- the delta's range between 1,15 and 2,23

- pasta's from different brands can be relatively similar, different brands of candy bars and different brands of instant soup are the most different



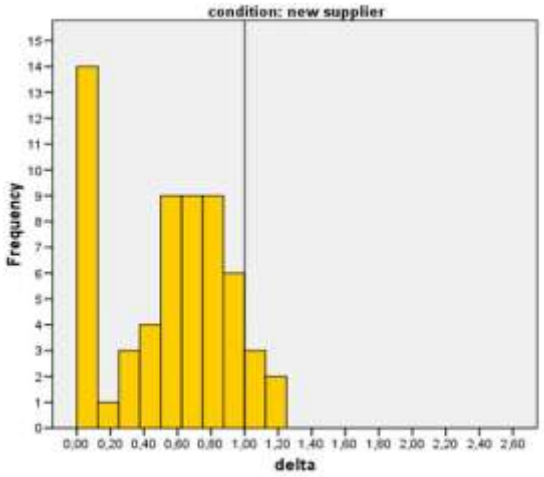


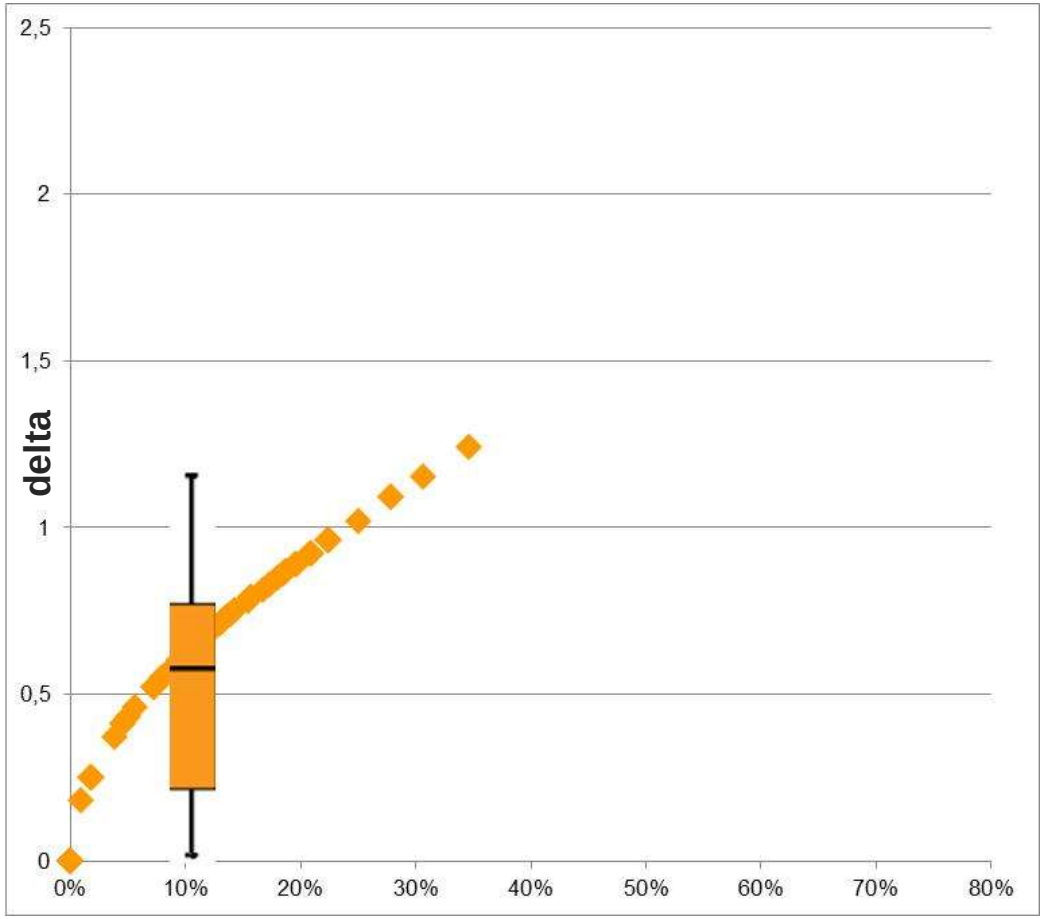
Delta's for supplier change (n=60)

can we make the same products?

- the median value of delta is 0,61
- the delta's range between 0 and 1,24

- on average, the delta is similar to the delta for dummies but their is more variability
- still, in 25% of the cases the two variants are indistinguishable (delta=0) but the match is not always completely successful



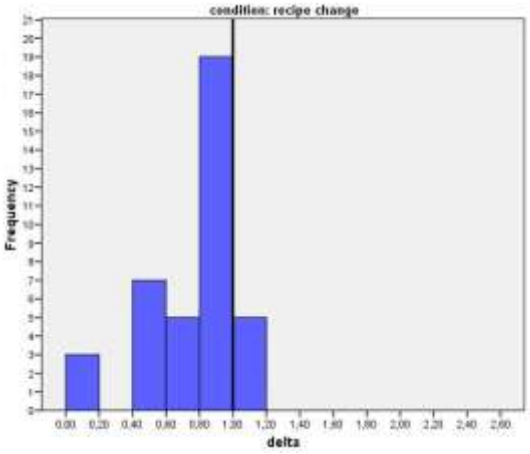


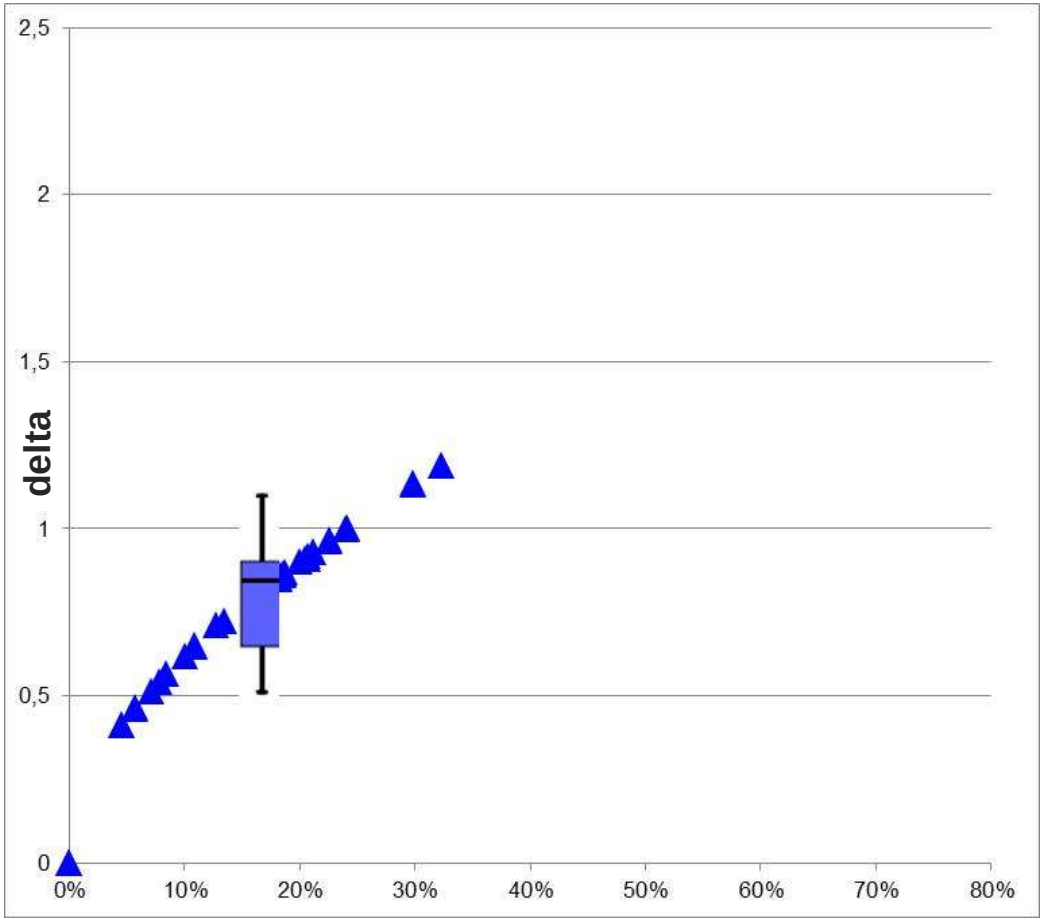
Delta's for new formulations (n=39)

how close can we match?

- the median value of delta is 0,85
- the delta's range between 0 and 1,19

- as can be expected, new formulations or recipe changes will or can not always be identical so delta is higher in this case than when switching from suppliers





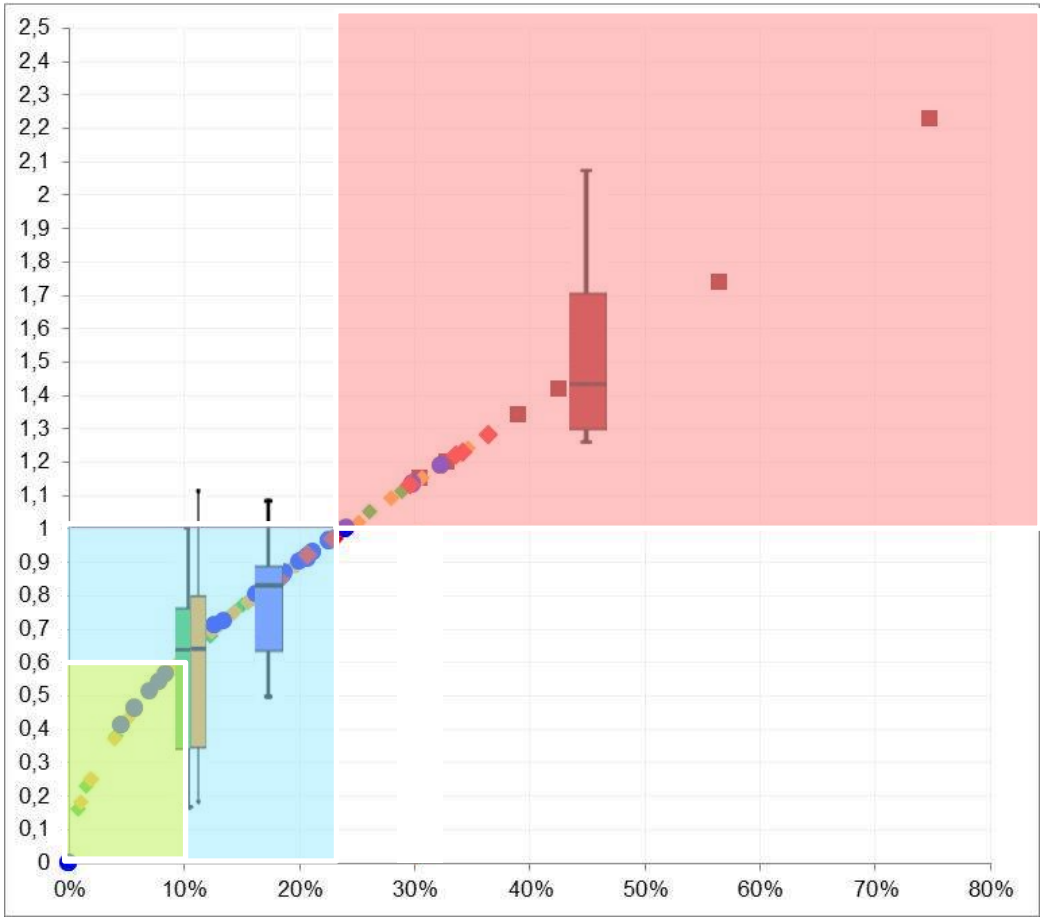
Defining benchmarks for consumer tetrad testing

- based on these results, the following benchmarks have been defined:

d' ≤ 0,60 means identical

0,60 < d' ≥ 1,00 means borderline

d' > 1,00 means NOT similar



Conclusion and some practical considerations

- consumers can perform tetrad tests with simple and complex products
- there can be considerable variation between ‘identical’ products and ‘different’ products can be quite similar
- still, meaningful benchmarks can be defined for tetrad tests
- there are no indications of disturbing sensory fatigue or problems with the task

Practical considerations

- potential problems with the protocol
  - fresh ingredients in soups
  - fresh ingredients in meals and pastas
  - cooking time in pasta
  - portioning
- potential sources of variation:
  - variability between sachets, production dates
  - subtle differences in preparation (cooking time, temperature)
  - differences in added ingredients (fresh meat, veggies, sauces)

who told you that difference testing was easy?

Thank you for your attention

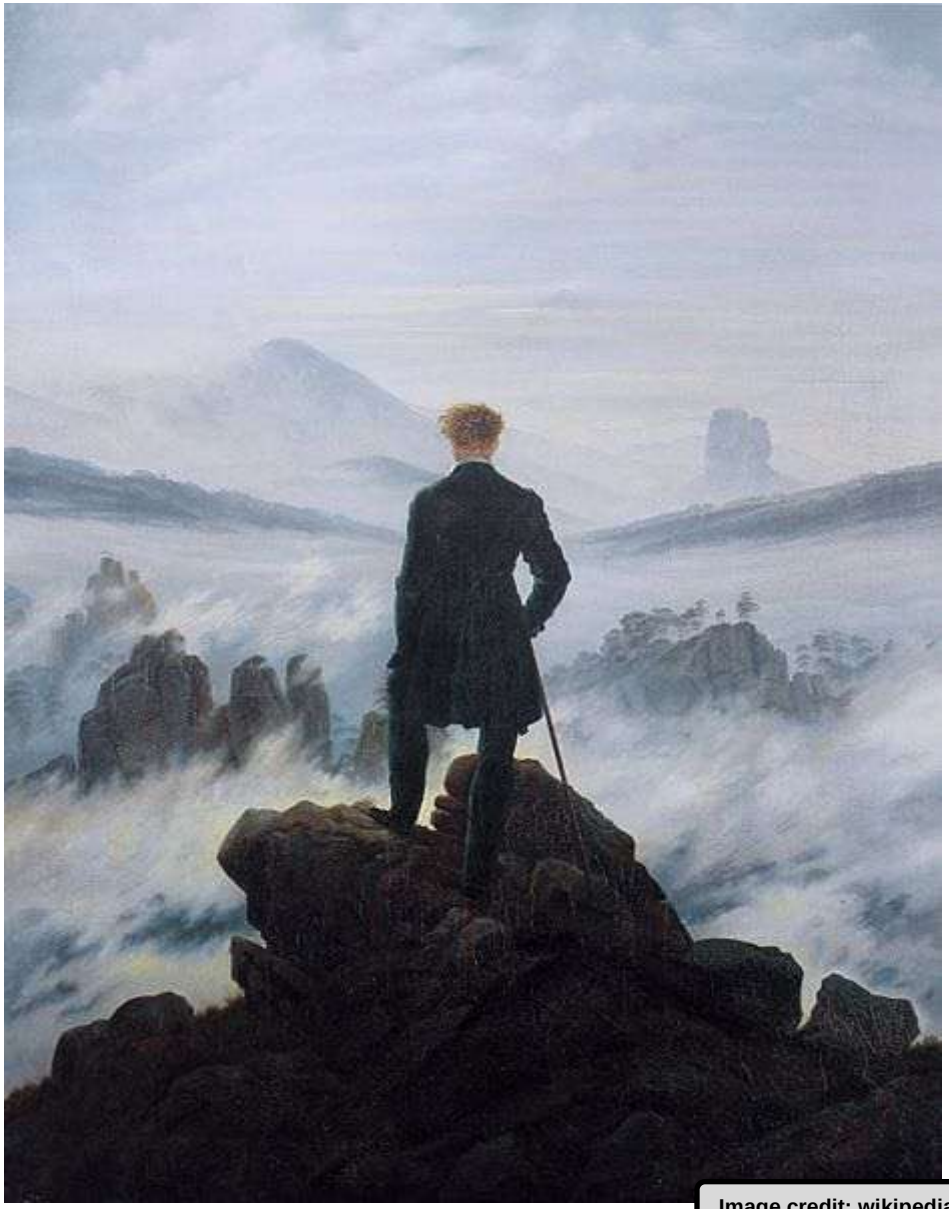


Image credit: wikipedia.org

Karen Garcia

Sensory Scientist at Symrise

- B.S. in Chemical Engineering
- M.S. and Ph.D. in Food Science from Louisiana State University
- Dissertation research focused on discrimination testing with children
- Conducted first large-scale comparison of Tetrad and Triangle testing

Thinking Inside the Box

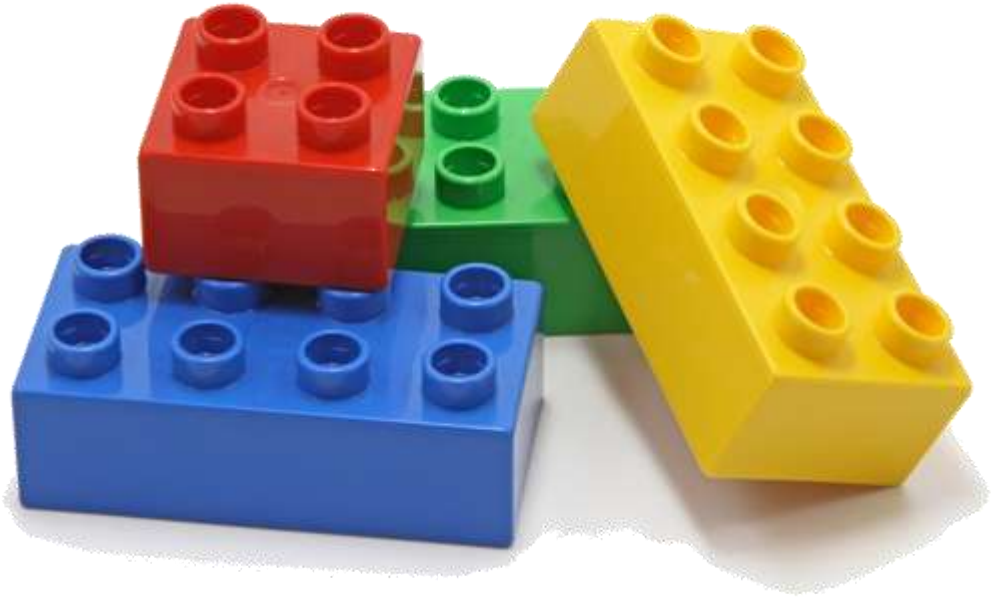
A Large-Scale Experimental Comparison of the Tetrad and Triangle Tests in Children

Karen Garcia – Louisiana State University (kgarci2@tigers.lsu.edu)
John Ennis – The Institute for Perception
Daniel Ennis – The Institute for Perception
Witoon Prinyawiwatkul – Louisiana State University

OBJECTIVE

- Comparison of the Tetrad and Triangle tests in a large scale study with children as subjects
  - Determine effect sizes of both methods
  - Validate the predicted higher power of the Tetrad over the Triangle test

MATERIALS AND METHODS



- **Subjects**
  - N = 404
  - 6-11 yoa

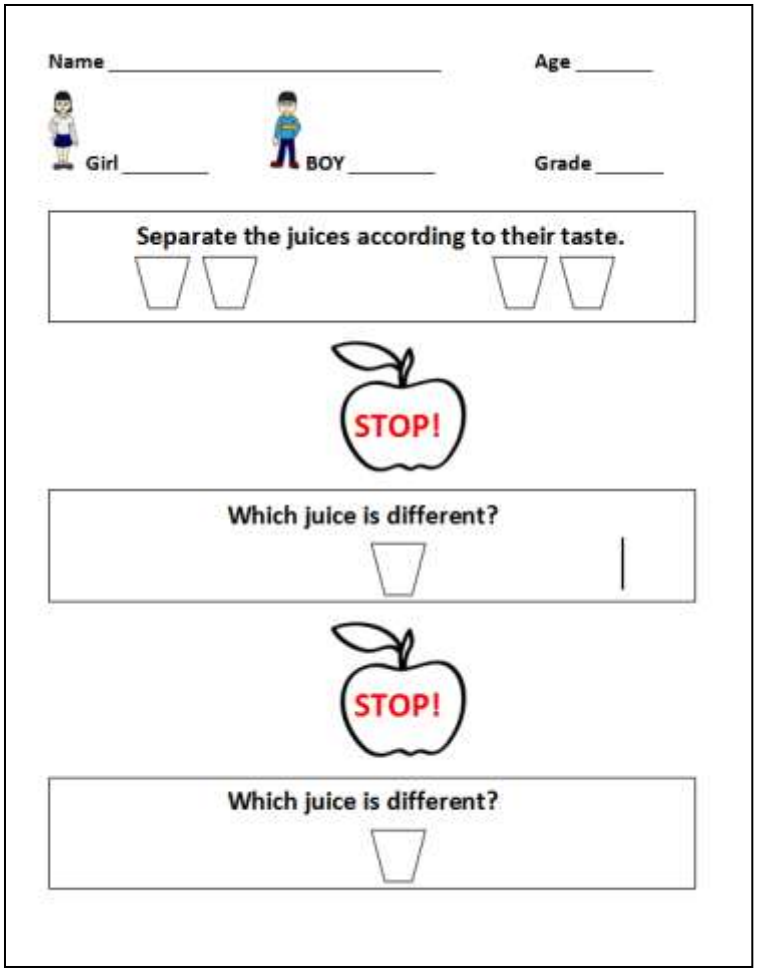
- **Stimuli**
  - 100% vs. 75% apple juice

- **Testing Procedure**
  - Parental Consent
  - Child Assent
  - Sample presentations
  - Tasting procedure demonstration
  - Questionnaire usage explanation

  - Tetrad instructions
    - *"Here are four juice samples; two belong to one group and the other two belong to a different group:* ***separate them according to their taste into two groups of two****"*

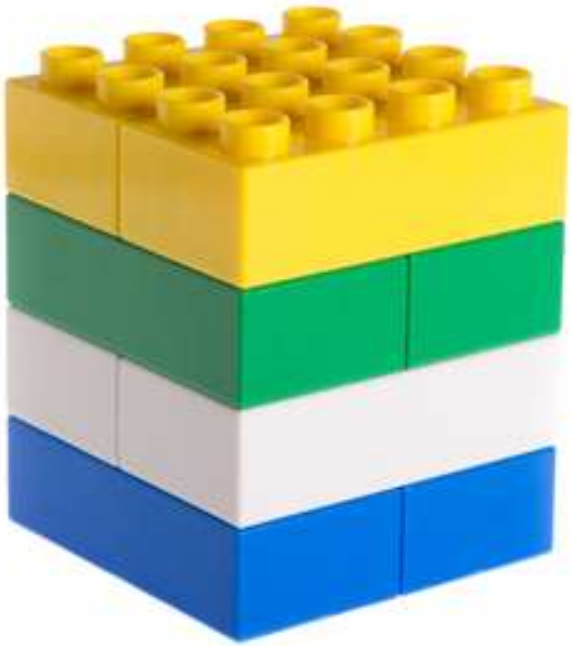
  - Triangle instructions
    - *"Here are three juices; two are the same and one is different:* ***identity the juice that is different****"*

- **Testing Procedure**
  - One trial Tetrad test
  - Two trials Triangle test

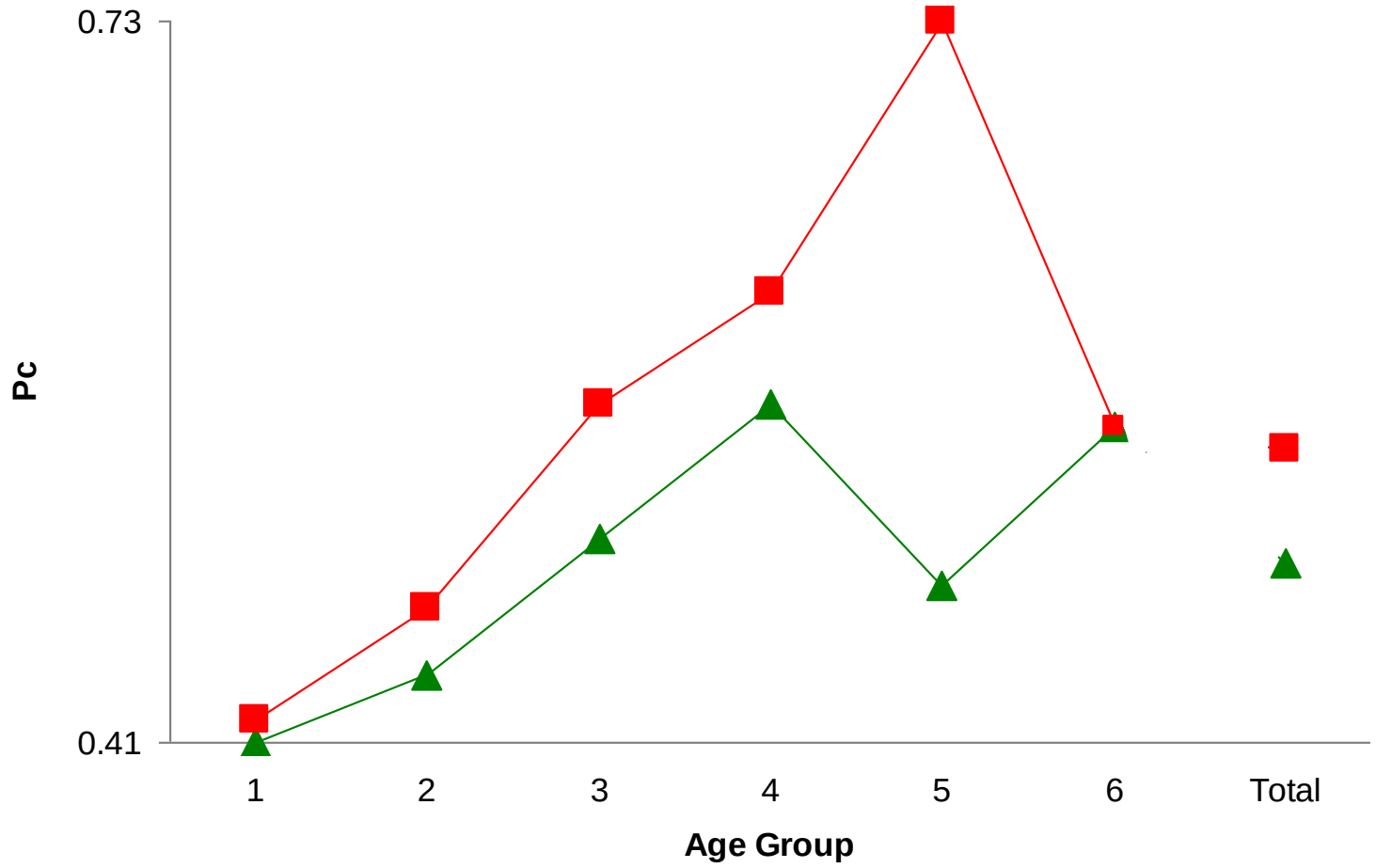


- **Data Analysis**
  - Effect sizes
    - Thurstonian theory (Ennis et al., 2012)
  - Perceptual noise and decreased effect sizes
    - Heuristic proposed by Ennis (2012)

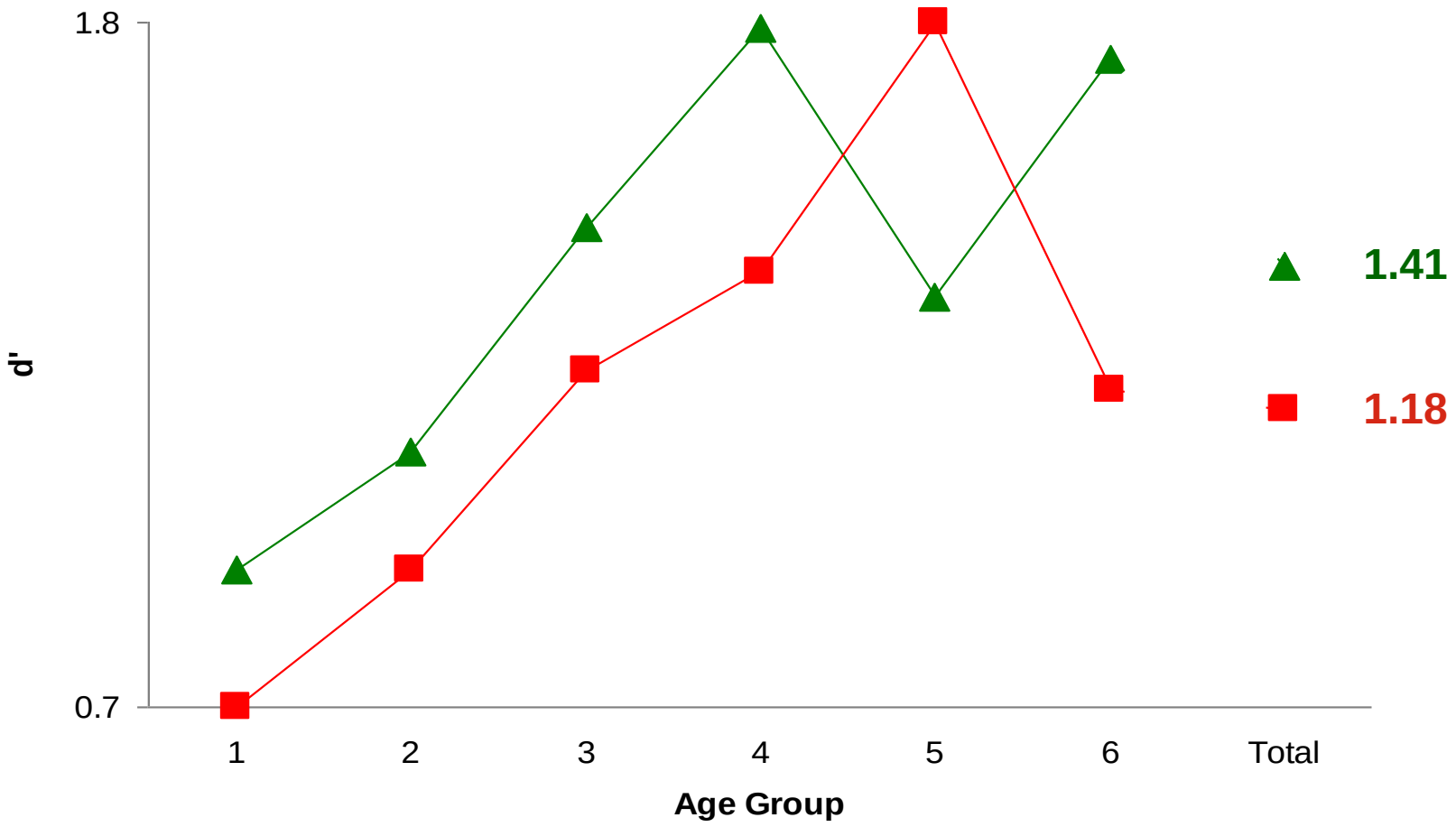
RESULTS



PROPORTION OF CORRECT RESPONSES



EFFECT SIZES (d’)



WAS THERE A DIFFERENCE THE EXPERIMENT COULD NOT DETECT?

- Sensory fatigue, adaptation, and memory effects could have led to the decreased effect size when a fourth stimulus was introduced
  - Special population

- Is there more perceptual noise in the Tetrad test?

- If so, how much additional perceptual noise allows the Tetrad test to still be more powerful than the Triangle test?

PERCEPTUAL NOISE EFFECT SIZES

- Heuristic proposed by Ennis (2012)
  - Effect size = signal-to-noise ratio
- Tetrad test is more powerful than the Triangle test as long as the introduction of the fourth stimulus
  - does not increase the perceptual noise by more than 50%
  - results in an effect size greater than 2/3 the triangle effect size
- Perceptual noise increase in Tetrad test of ~20%
- $\delta_{\square}=1.18,\ \delta_{\Delta}=1.41$
- $\delta_{\square} > {}^{2}/_{3}\ \delta_{\Delta}$ satisfied

CONCLUSION

- Need for a large-scale test comparing the Tetrad and Triangle tests satisfied
- For sweetened apple juice and children as subjects
  - Tetrad test had a higher proportion of correct responses vs. Triangle
  - Tetrad test had reduced effect sizes vs. Triangle
- Tetrad test remained superior than the Triangle test in this setting

Karen Garcia – Louisiana State University (kgarci2@tigers.lsu.edu)
John Ennis – The Institute for Perception
Daniel Ennis – The Institute for Perception
Witoon Prinyawiwatkul – Louisiana State University

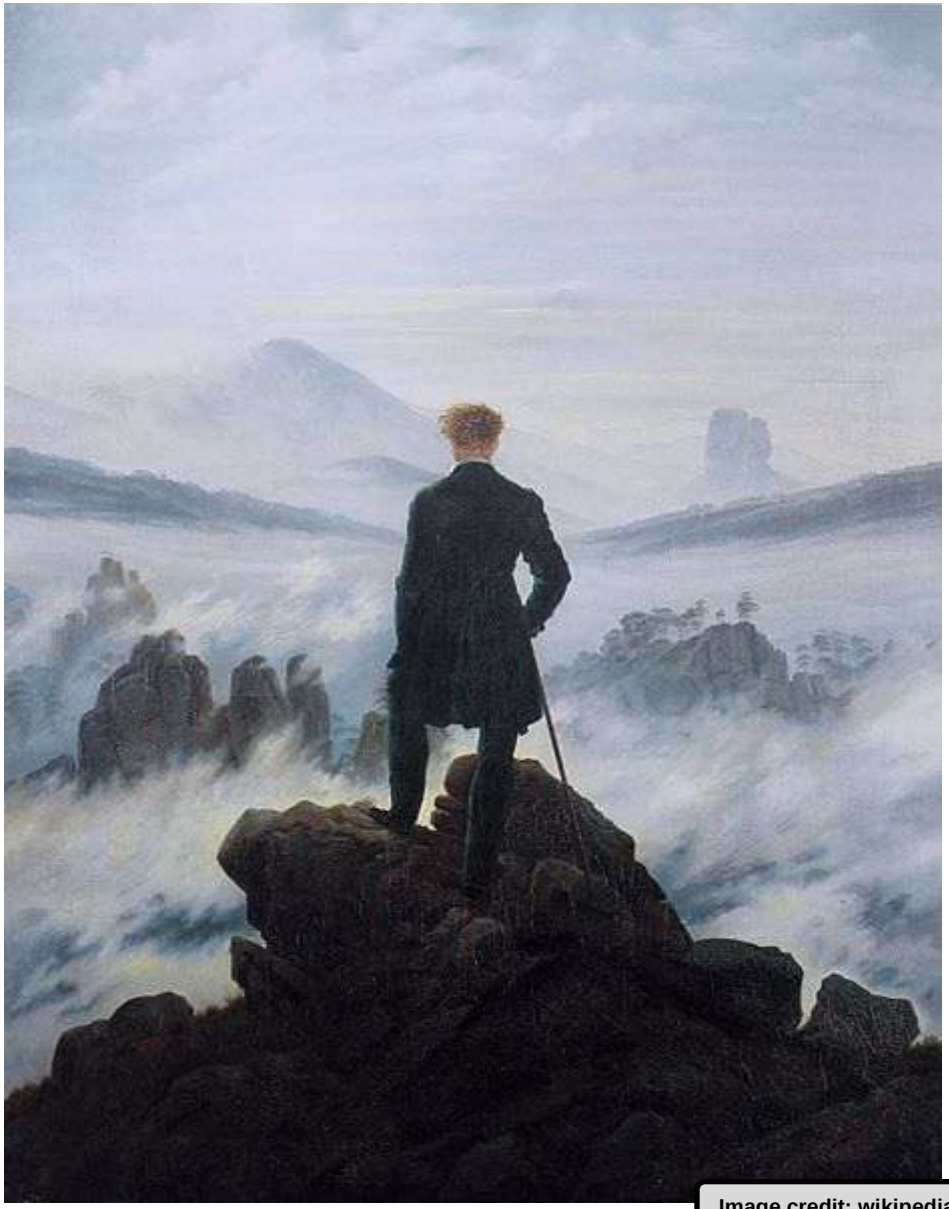


Image credit: wikipedia.org

John M. Ennis

Vice President of Research Operations at
The Institute for Perception

- Ph.D. in Mathematics from University of California, Santa Barbara
- Post-doctoral studies in Psychology
- Extensive project management experience for international food and personal care product companies
- Publications and presentations in Sensory Science, Market Research, Statistics, Mathematics, and Psychology
- Co-author of “Short Stories in Sensory and Consumer Science”
- Chair of ASTM subcommittee E18.04 - “Fundamentals of Sensory”

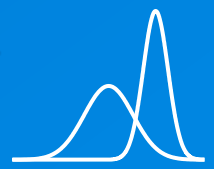
Society of Sensory Professionals
Jersey City, New Jersey
October 11th, 2012

The Sensitivity of the Tetrad, Triangle, and Degree of Difference Tests

John Ennis – The Institute for Perception
Rune H.B. Christensen – Technical University of Denmark
Daniel Ennis – The Institute for Perception
Benoit Rousseau – The Institute for Perception

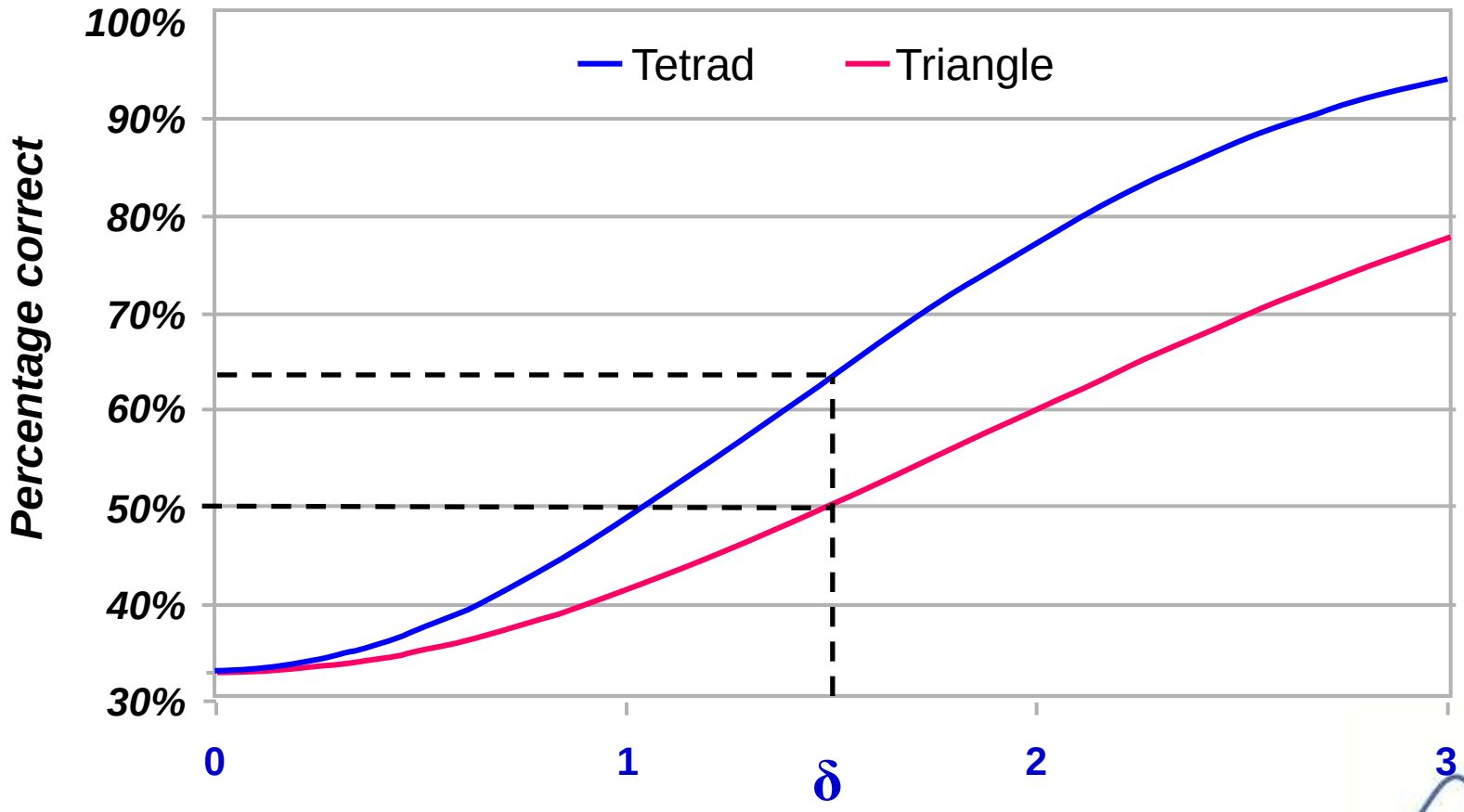
john.m.ennis@ifpress.com **rhbc@imm.dtu.dk**

www.ifpress.com

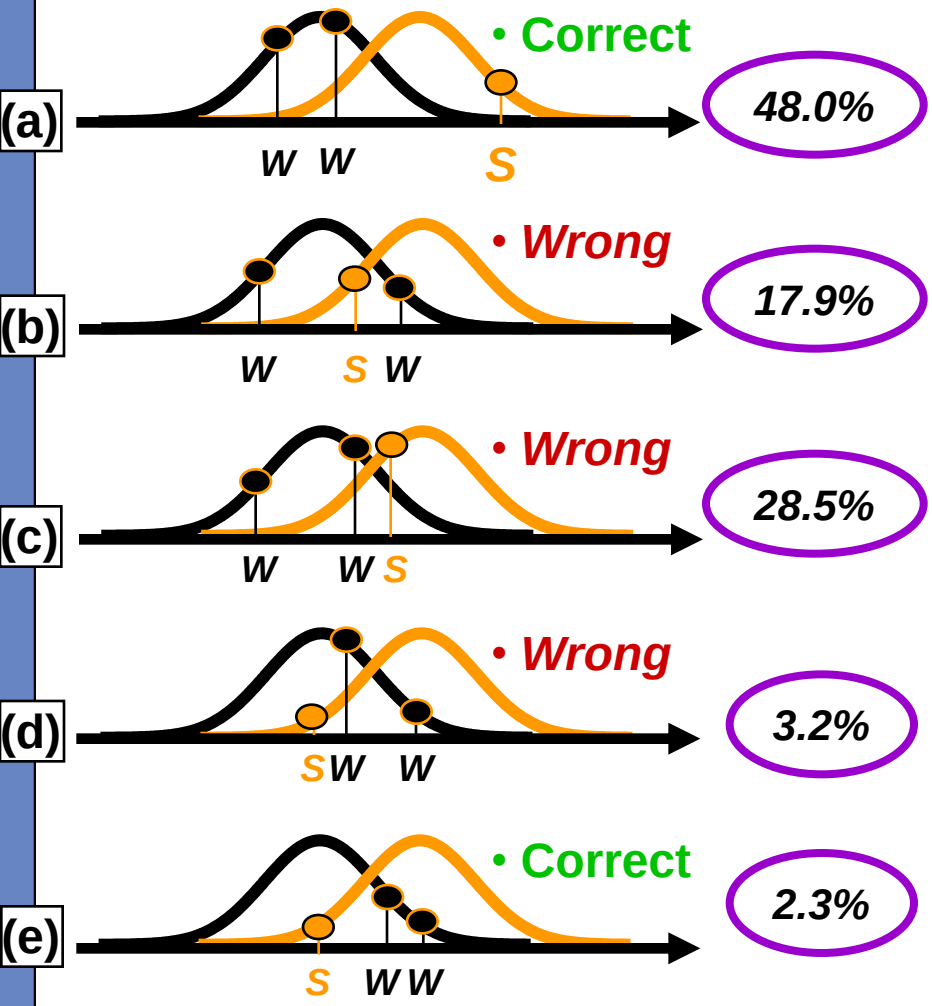


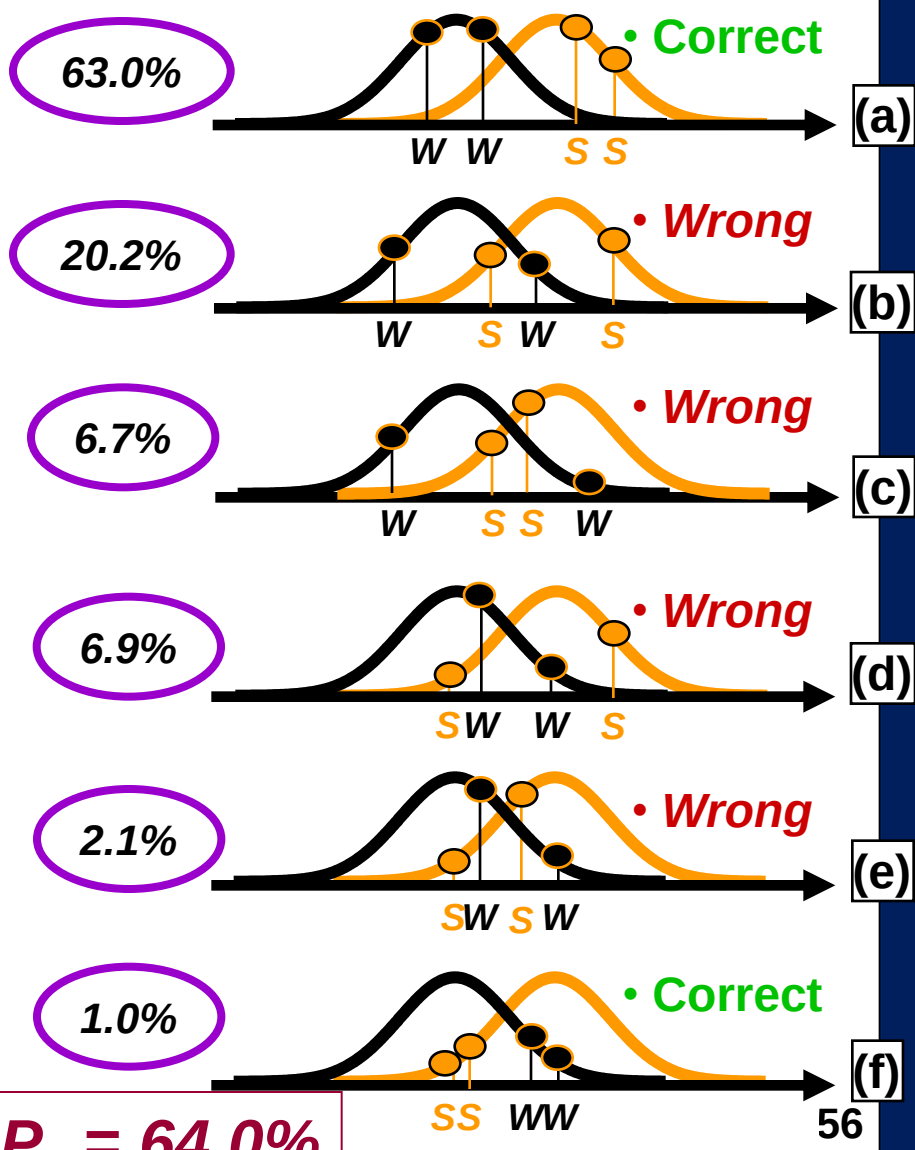
Estimating Sensory Differences

- As sensory difference increases, both Triangle and Tetrad give more correct answers
  - But Tetrad is more responsive



Triangle/Tetrad – Possible Cases (δ = 1.5)



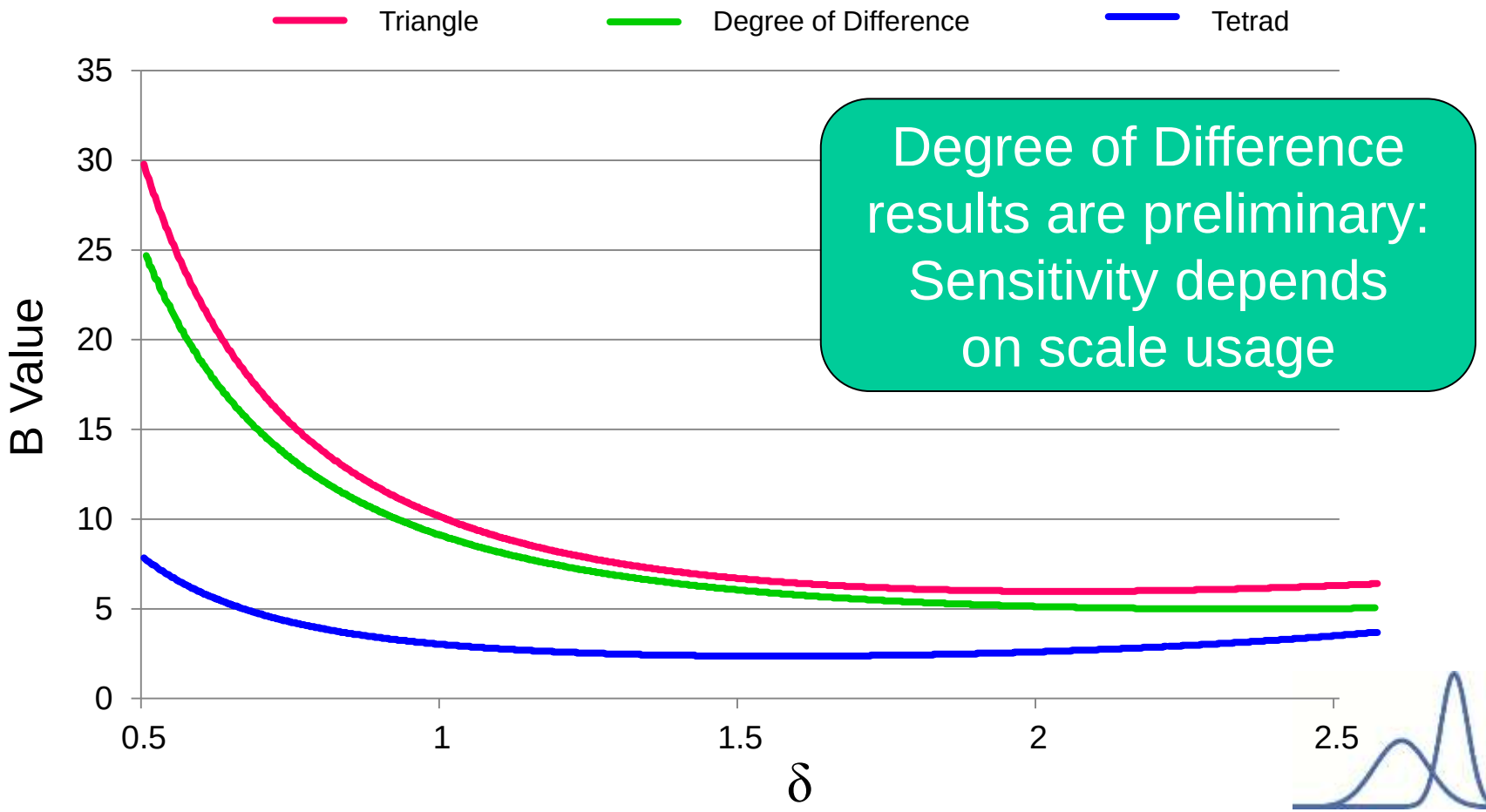


P_c = 50.3%

P_c = 64.0%

Precision of Measurement (1/2)

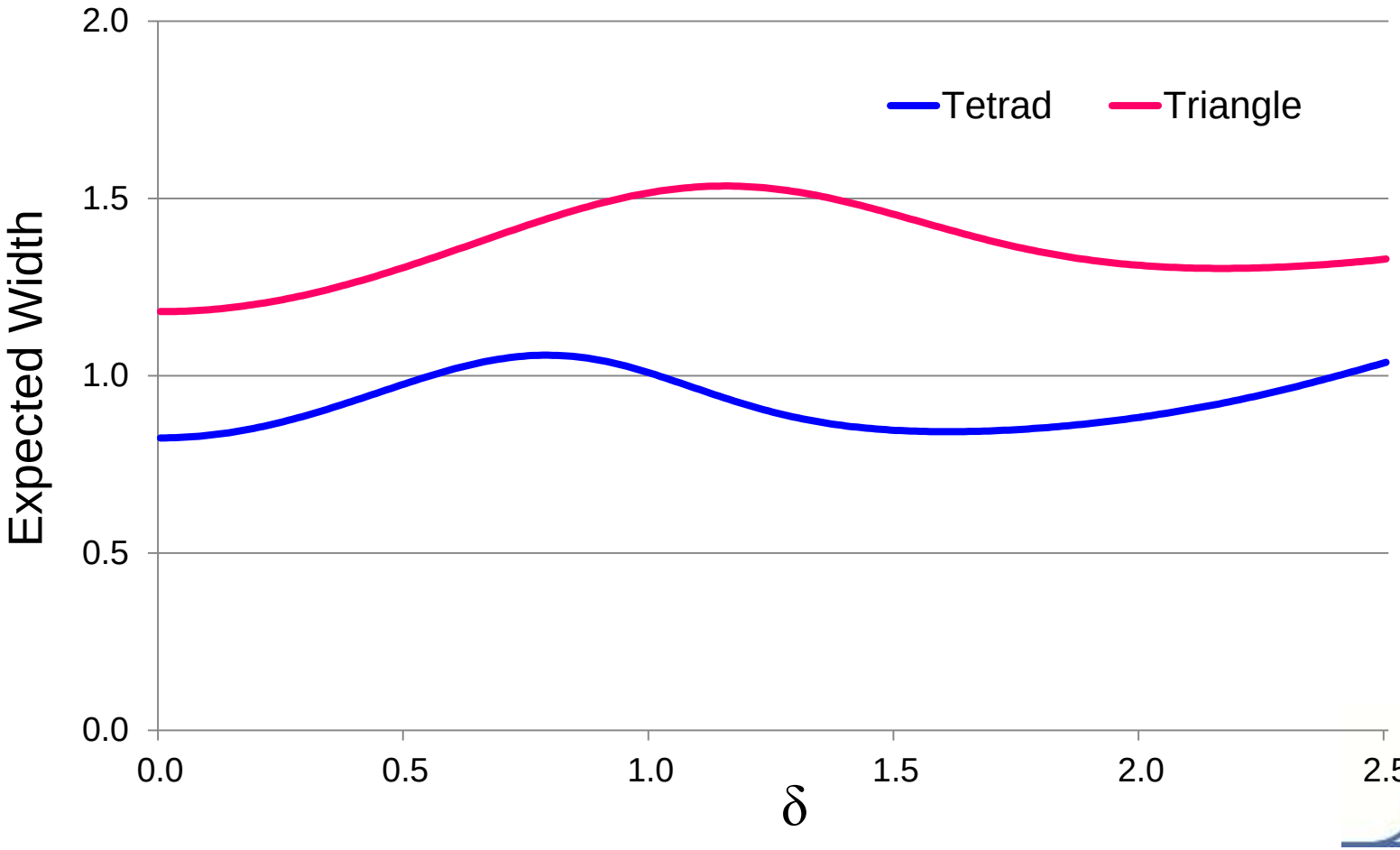
- Variance in estimate of δ (Bi, Ennis, & O'Mahony, 1997)
  - Variance is B value divided by sample size



Degree of Difference results are preliminary: Sensitivity depends on scale usage

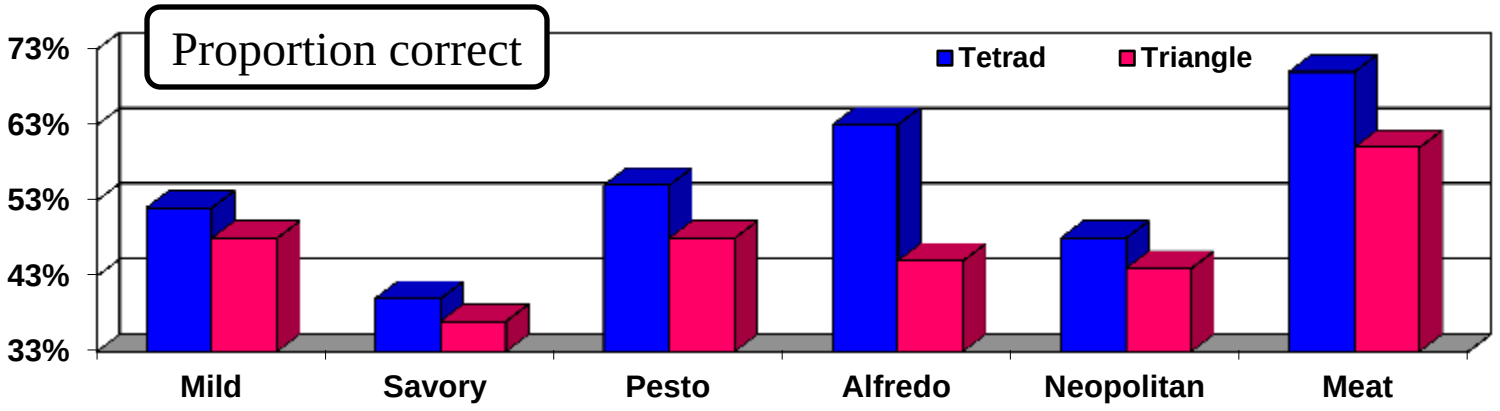
Precision of Measurement (2/2)

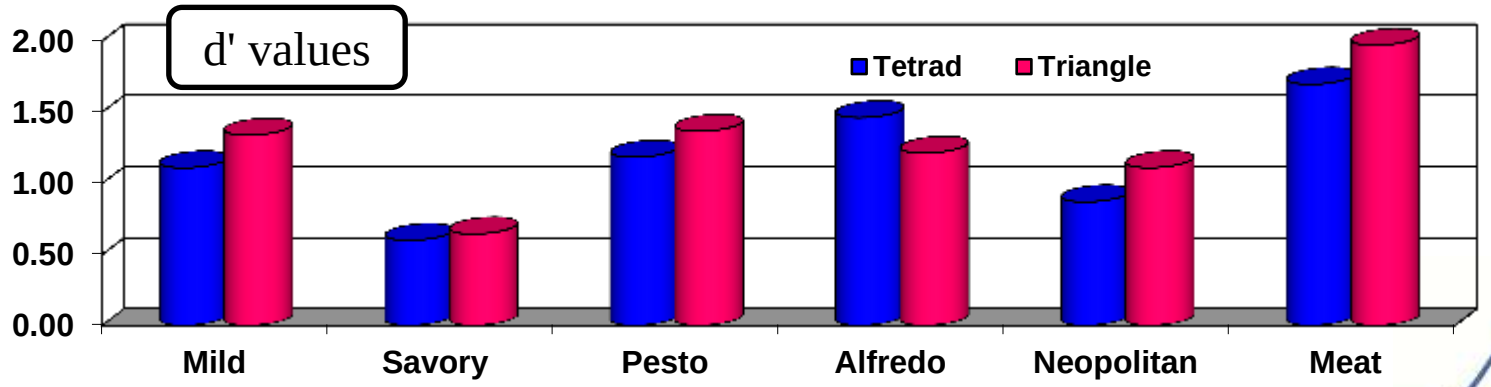
- Expected widths of likelihood confidence intervals
  - N = 60, 95% confidence



Comparative Examples (1/2)

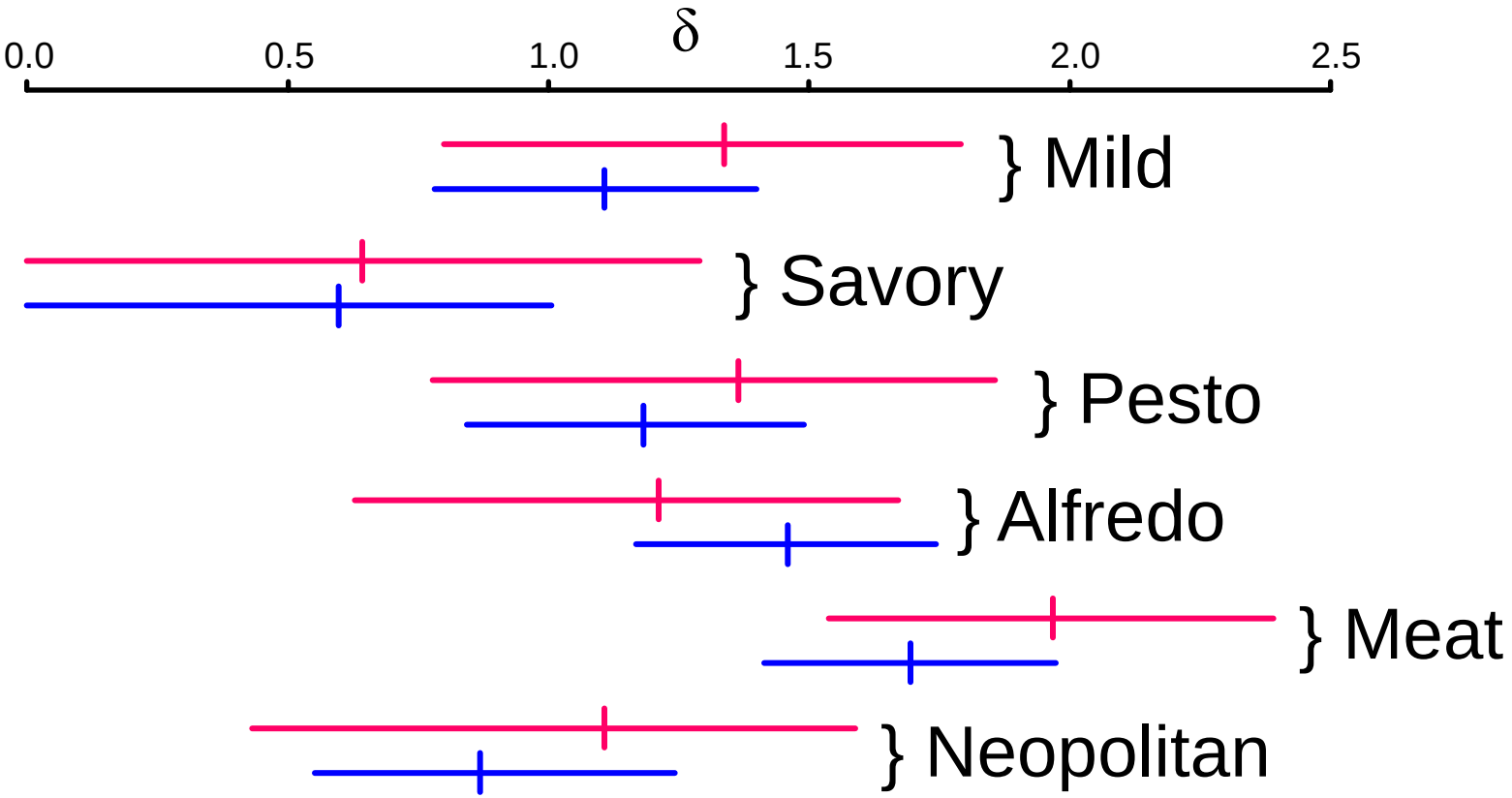
- Six pasta sauces for food service applications
- Research to compare Triangle and Tetrad tests
  - Test sample sizes vary between 96 and 132





Comparative Examples (2/2)

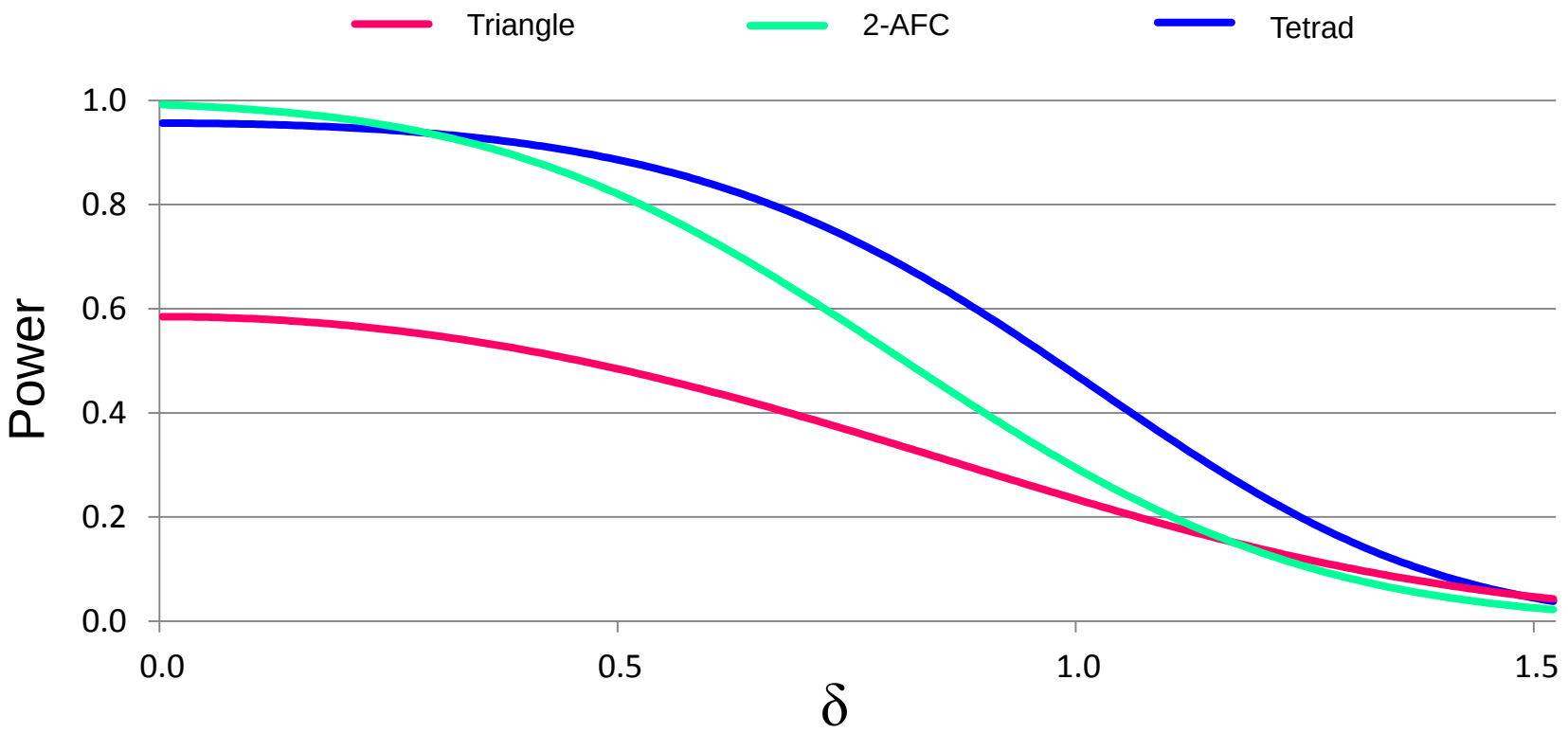
- Likelihood confidence intervals:



- Tetrad test gives more precise estimate of sensory difference in each case

Equivalence Testing

- Tetrad testing is also more powerful than Triangle testing for Equivalence testing (Ennis & Christensen, *in review*):



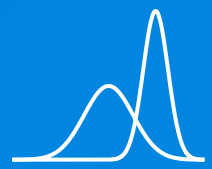
- Under some circumstances, Tetrad test is even more powerful than the 2-AFC

Society of Sensory Professionals
Jersey City, New Jersey
October 11th, 2012

The Sensitivity of the Tetrad, Triangle, and Degree of Difference Tests

John Ennis – The Institute for Perception
Rune H.B. Christensen – Technical University of Denmark
Daniel Ennis – The Institute for Perception
Benoit Rousseau – The Institute for Perception

john.m.ennis@ifpress.com



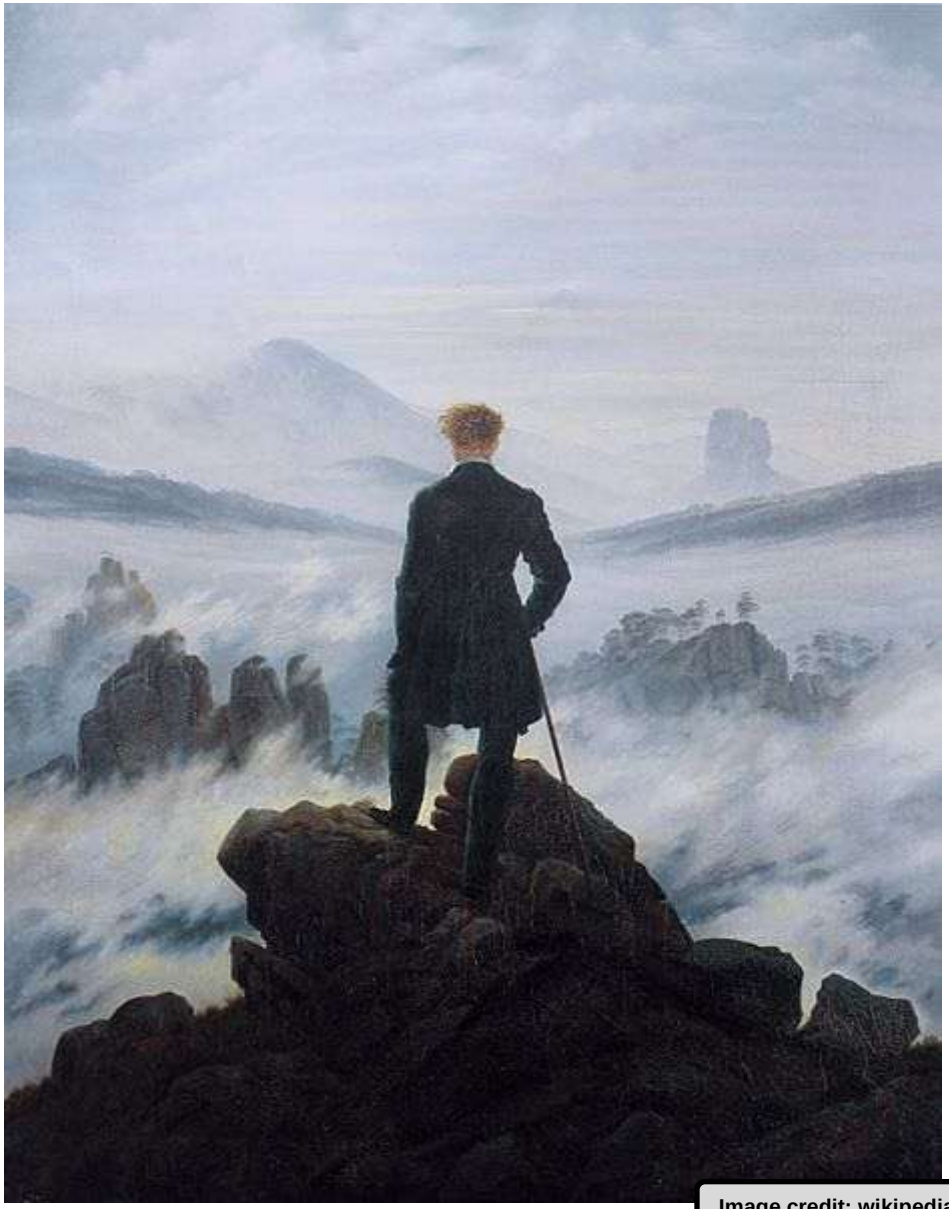


Image credit: wikipedia.org

Society of Sensory Professionals

Jersey City, New Jersey

October 11th, 2012

Technical Workshop – An Exploration of Tetrad Testing