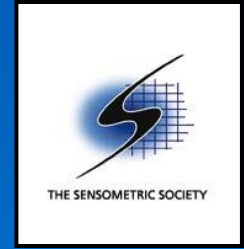




Sensometrics 2014
Chicago, Illinois, USA



A Thurstonian Comparison of the Tetrad and Degree of Difference Tests

John M. Ennis
Benoît Rousseau

The Institute for Perception

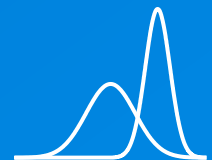
john.m.ennis@ifpress.com

benoit.rousseau@ifpress.com

Rune H.B. Christensen

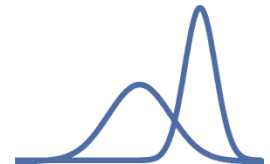
Technical University of Denmark

rhbc@imm.dtu.dk



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 of measurement
 3. Determine size of consumer relevant difference





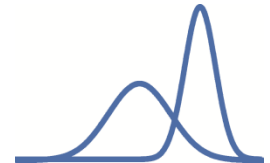
The Tetrad Test - Methodology

- Four samples presented:



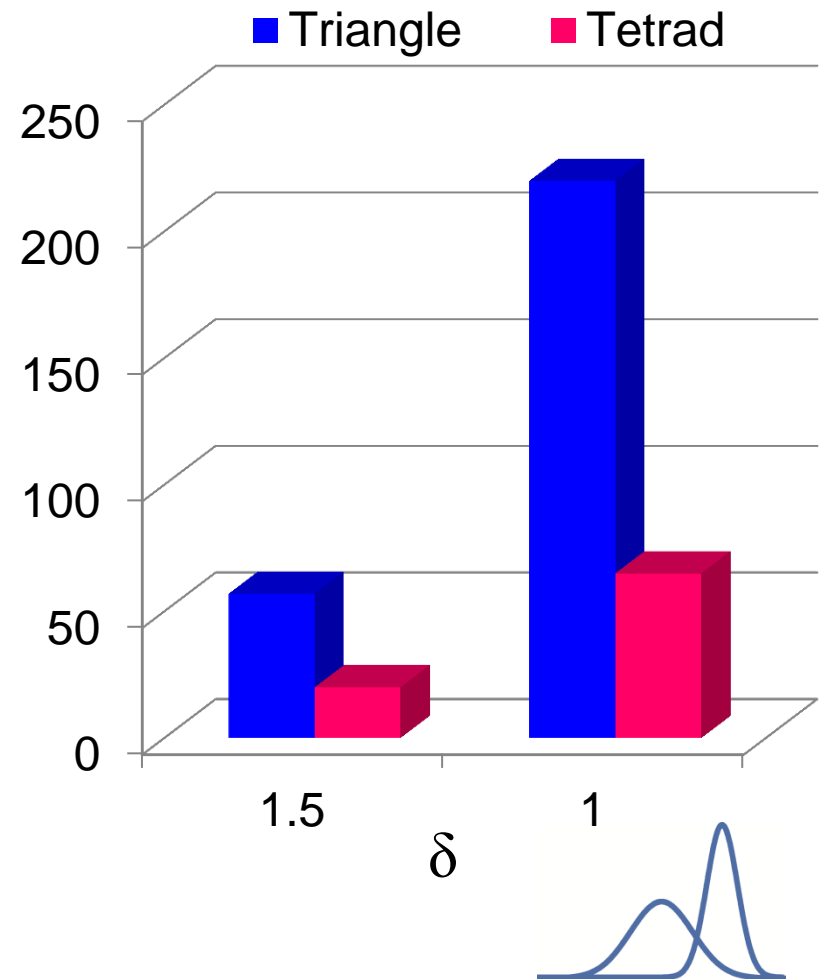
“Group the stimuli into two groups of two samples based on similarity”

- Six possible presentation orders: AAB^B, AB^AB, ABBA, BB^AA, BA^BA, BAAB
- Guessing probability = $1/3$



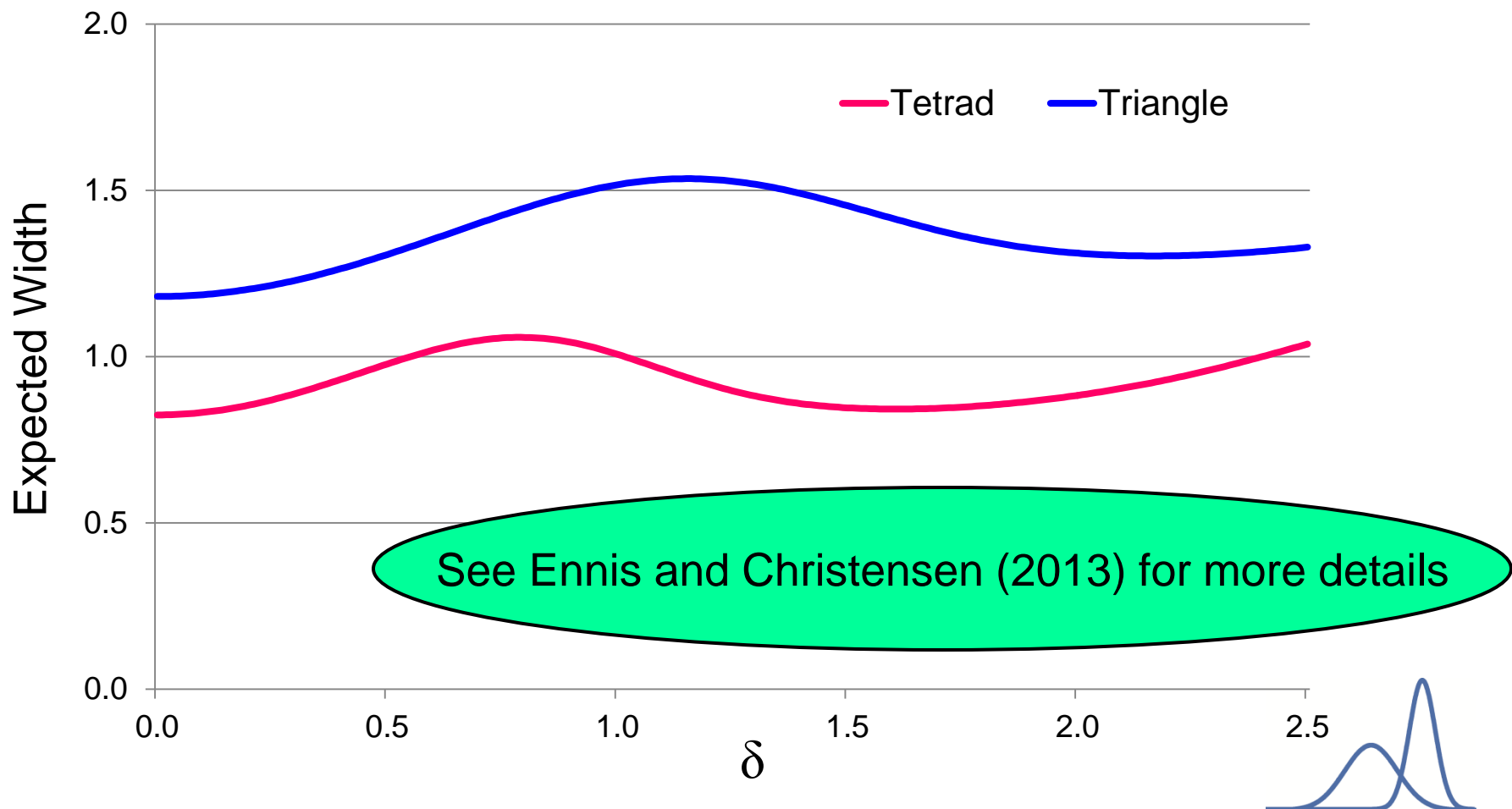
Tetrad/Triangle – Sample Sizes

- Suppose $\alpha = 0.05$ and want 80% power
- If $\delta = 1.5$
 - ❖ Triangle $N = 57$
 - ❖ Tetrad $N = 20$
- If $\delta = 1.0$
 - ❖ Triangle $N = 220$
 - ❖ Tetrad $N = 65$
- Under ideal conditions, sample sizes for Tetrad are roughly 1/3 sample sizes for Triangle
- See Ennis & Jesionka (2011) for more information



Triangle/Tetrad – Precision

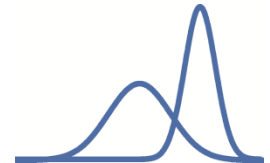
- Expected widths of likelihood confidence intervals
 - ❖ $N = 60$, 95% confidence



Practical Considerations



- Tetrad test uses four stimuli
 - ❖ Adding a stimulus may cause additional perceptual noise
- Even with additional noise, Tetrad will be more powerful than Triangle if additional noise $\leq 50\%$ (Ennis, 2012)
- Need to verify this result for product category of interest
 - ❖ See Ennis (2012) for more details
- Question: How does the Tetrad test compare to other difference tests that use fewer stimuli?
 - ❖ The Degree of Difference (DoD) test only requires two stimuli is in common use
 - ❖ Even if Tetrad is theoretically superior to DoD, how much additional noise can be allowed before advantage is lost?



Degree of Difference

- There are many versions of “Degree of Difference”
 - ❖ We consider rated Same/Different task



*Sample
1*



*Sample
2*

Instructions

*How different are the two samples
on a scale from 1 to N?*

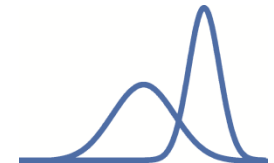


Extremely Similar



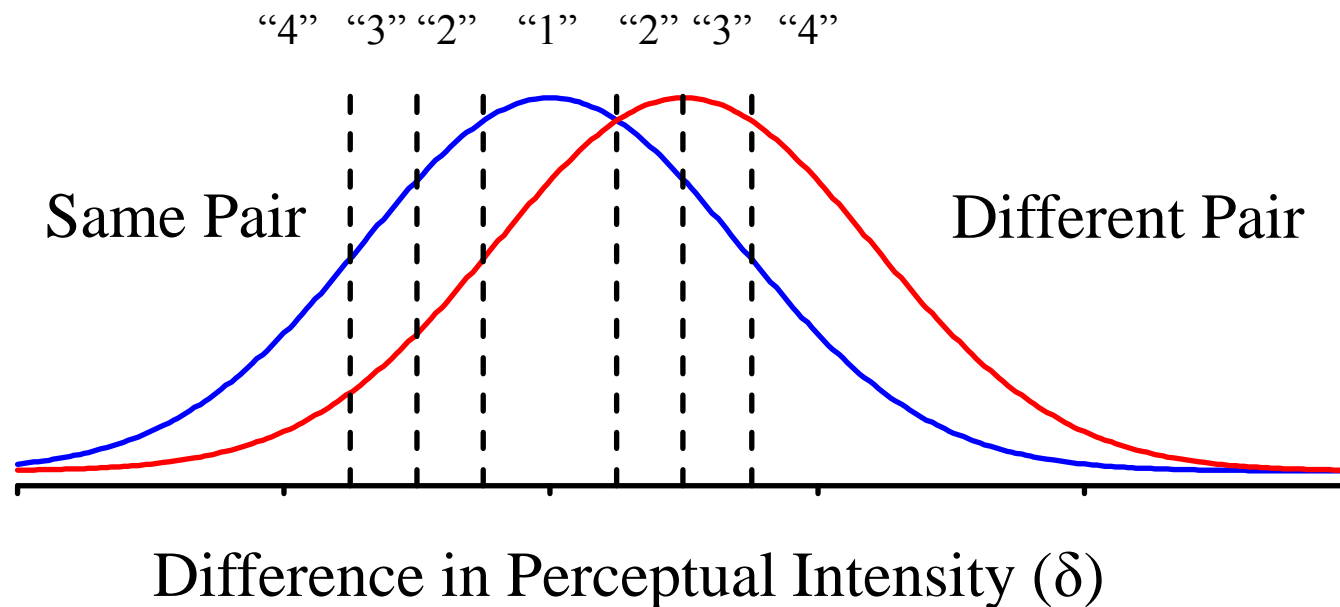
*Extremely
different*

- Useful when evaluation of more than 2 samples is difficult
- Notes:
 - ❖ When $N = 2$, have standard Same/Different task
 - ❖ Fu and Rousseau (2011) argued for 4 categories
 - ❖ Typically equal numbers of same and different pairs

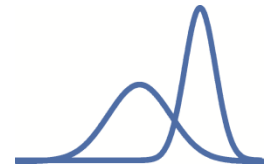


Thurstonian Model for DoD

- Ennis (1993) proposed a Thurstonian model for DoD
 - ❖ Considers difference distributions



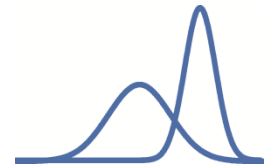
- Factors affecting performance (Ennis and Christensen, 2014)
 - ❖ Number of categories
 - ❖ Ratio of same pairs to different pairs
 - ❖ Placement of scale boundaries



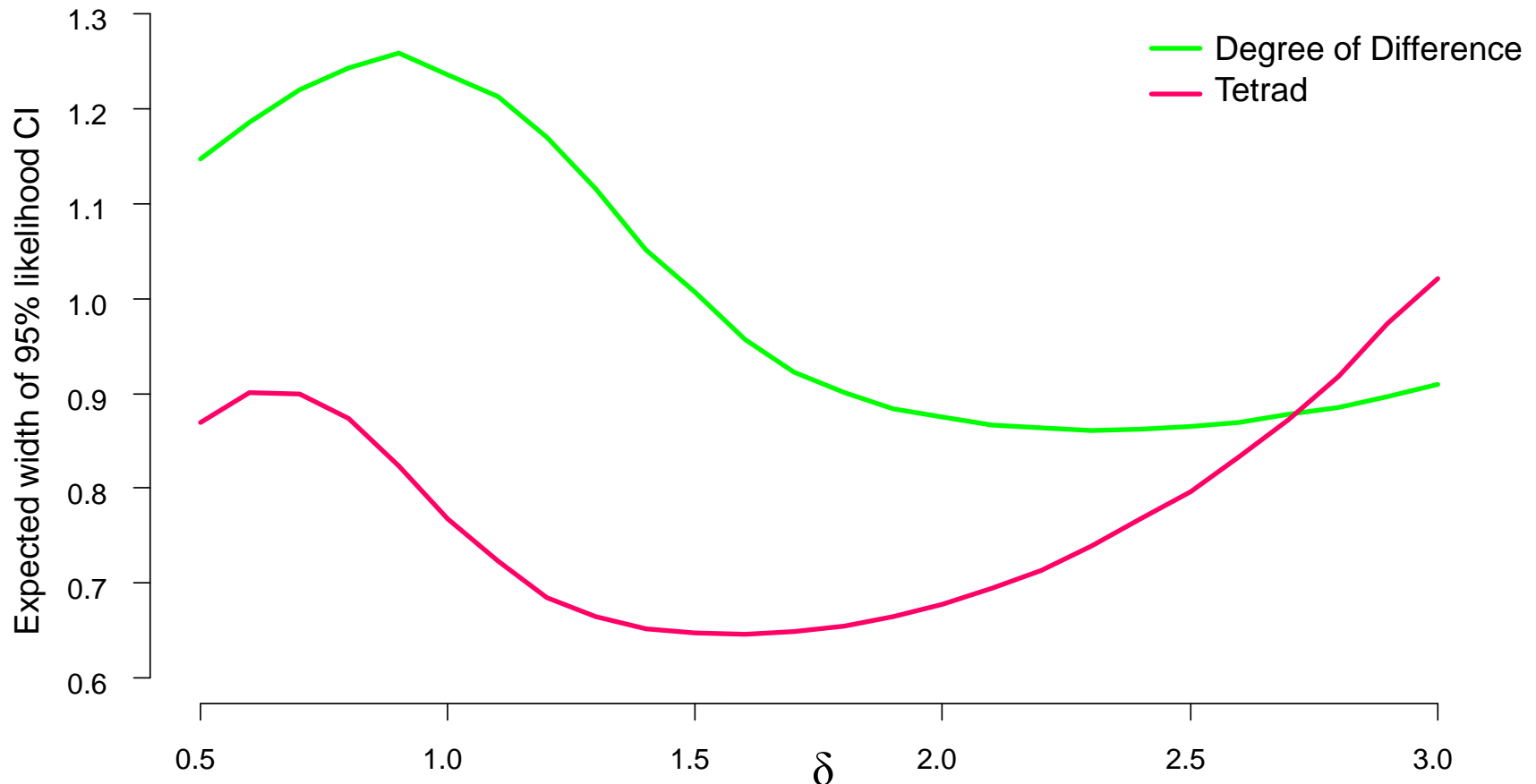


Comparative Analyses

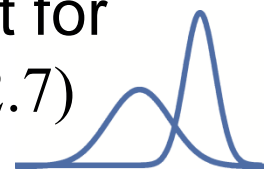
- With the parameters of the DoD optimized, the Tetrad and DoD were compared with respect to precision and power
- Precision:
 - ❖ The expected widths of the 95% profile likelihood confidence intervals were compared
- Power:
 - ❖ 100,000 simulated experiments for 25 equally spaced values of δ between 0 and 3 were considered, for a variety of sample sizes
 - ❖ A Wilcoxon rank sum test was applied in each simulated experiment
 - ❖ The estimated power of the DoD was then compared to the known power of the Tetrad



Precision Comparison ($N = 100$)

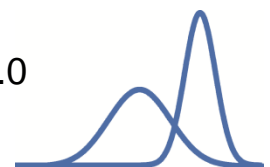
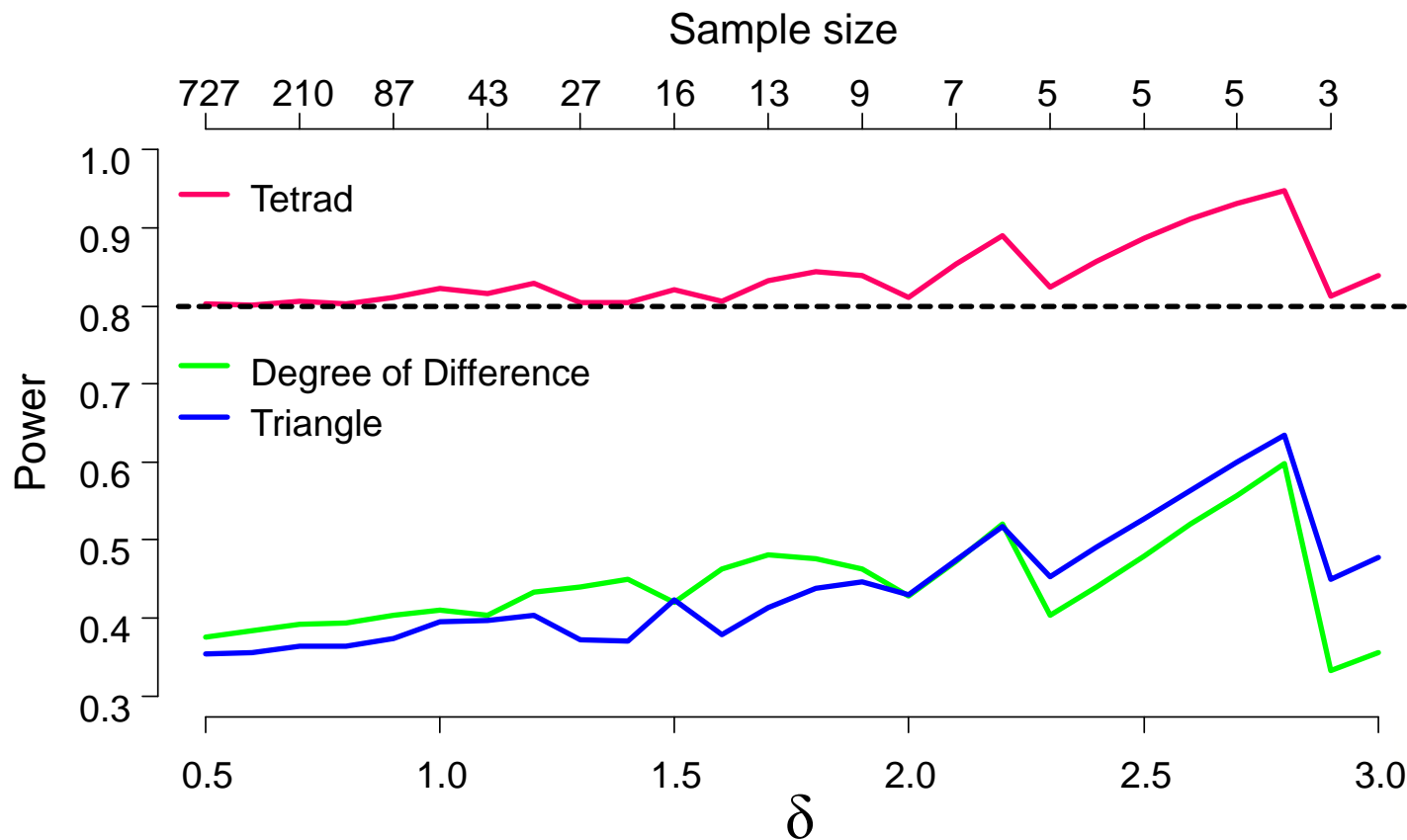


- The Tetrad test is more precise than the DoD test for all δ values likely to be of practical interest ($\delta \leq \sim 2.7$)



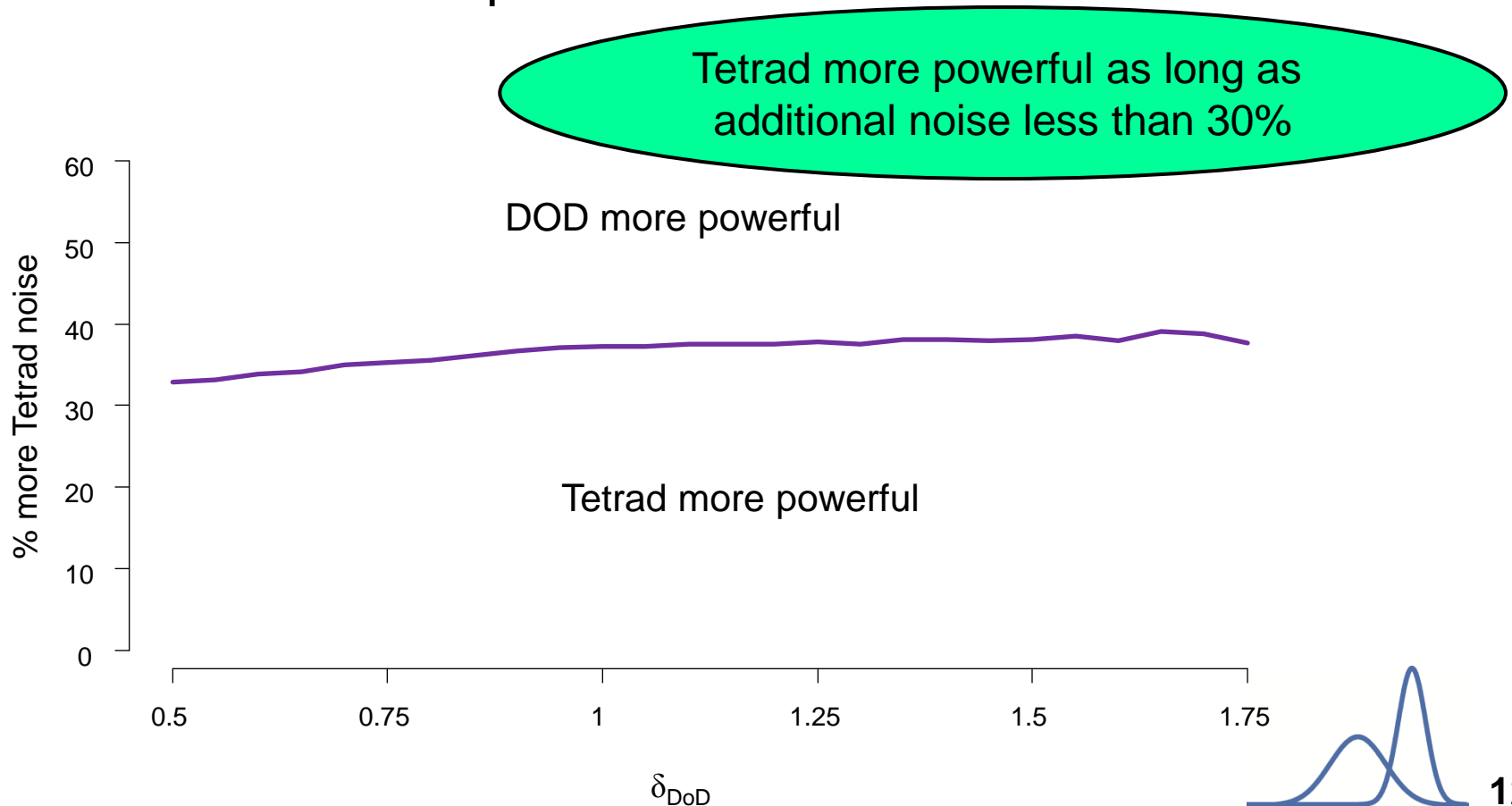
Power Comparison

- To compare power across sample sizes, the power of the DoD for the minimum sample sizes that give 80% power for Tetrad was estimated



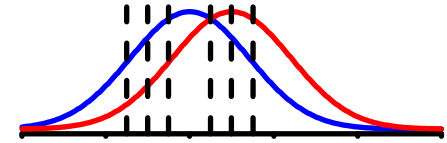
Isopower Comparison

- The amount of additional noise the Tetrad test can withstand before losing its power advantage over the DoD was also computed

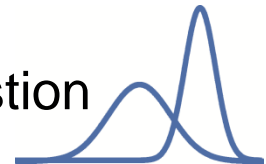




Summary

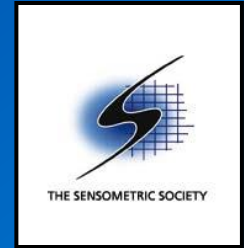


- The Tetrad and Degree of Difference (DoD) tests can be compared using Thurstonian theory
- Once optimal parameters are chosen for the DoD, and under ideal circumstances
 - ❖ The Tetrad test is substantially more precise
 - ❖ The power of the DoD is slightly greater than the Triangle and substantially lower than the Tetrad
- The Tetrad test requires four stimuli while the DoD only requires two
 - ❖ Additional perceptual noise from the additional stimuli could hurt the Tetrad test performance
 - ❖ The power of the Tetrad will be greater than the power of the DoD as long as additional noise does not exceed 30%
 - ❖ Whether or not this is the case is an experimental question





Sensometrics 2014 Chicago, Illinois, USA



Thank You!

**John M. Ennis
Benoît Rousseau**

The Institute for Perception

john.m.ennis@ifpress.com

benoit.rousseau@ifpress.com

Rune H.B. Christensen

Technical University of Denmark

rhbc@imm.dtu.dk

