



**10th Pangborn Sensory
Science Symposium**
Rio de Janeiro, Brazil
August 15th, 2013



***Developments in Applicability and
CATA scoring***

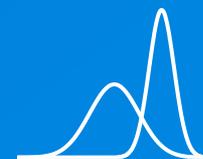
John M. Ennis

Daniel M. Ennis

The Institute for Perception

john.m.ennis@ifpress.com

daniel.m.ennis@ifpress.com



History of Applicability Scoring (1/2)

- In 1982, Loh and Ennis constructed a monadic testing instrument to measure the applicability of 36 phrases to a product
 - ❖ Responses were binary – “Applies” and “Does Not Apply”
 - ❖ One attribute was of key interest, remaining 35 attributes were distractors
- Goal: Determine whether a difference on the key attribute was noticeable to consumers
 - ❖ Many of the phrases were not suited to a paired comparison or rating format
- Approaches compared
 - ❖ 7-point rating scales and a 2-AC preference question
 - ❖ “Glossary of Attributes”, later called “Applicability Scoring”



History of Applicability Scoring (2/2)

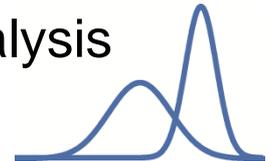
➤ Glossary of Attributes:

“Read the phrases and, for each phrase, mark the box to the left if the phrase describes the product that you just sampled. Mark the box on the right if the phrase does not describe the product tested.”

Attribute	Applies	Does Not Apply
Tastes harsh	<input type="checkbox"/>	<input type="checkbox"/>
Has a dense rod	<input type="checkbox"/>	<input type="checkbox"/>
Has a light taste	<input type="checkbox"/>	<input type="checkbox"/>
⋮	⋮	⋮

➤ Results

- ❖ Applicability scoring more sensitive than ratings
- ❖ Richer source of data than ratings based on factor analysis





Practical Value of Applicability Scoring

Loh and Ennis identified many benefits to applicability scoring, including:

- Quantitative information obtained on attributes not easily posed in rating or paired comparison format
- Easy for subjects to use with less demand
- Liking or derived preference
 - ❖ Applicability of “I like this product”
 - ❖ Can distinguish between *like both* and *like neither* which is not obtained from *no preference* in paired tests
- Found to be highly sensitive to product preferences



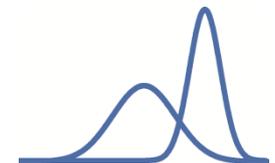


History of Forced-Choice CATA

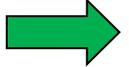
- Long history of CATA investigations in survey literature
 - ❖ e.g. Smyth et al. (2006)
- Sudman & Bradbury (1982): Forced-Choice CATA
 - ❖ Methodology same as Applicability Scoring
 - ❖ Used extensively for telephone surveys
- Checklist CATA commonly used for mail and internet

“Please check all of the phrases below that apply to the product that you just sampled.”

Tastes harsh	<input type="checkbox"/>
Has a dense rod	<input type="checkbox"/>
Has a light taste	<input type="checkbox"/>
⋮	⋮



Checklist vs Forced-Choice CATA

- “Satisficing” a potential problem with checklist CATA
 - ❖ Once a sufficient number of items have been selected, respondents may disengage from survey
 - ❖ Item appearing earlier in the list tend to be chosen more often
- Forced-Choice CATA leads to deeper processing - compared to checklist CATA:
 - ❖ Greater time used to completed survey
 - ❖ More items selected
- Key question for analysis: Unselected  “Does not apply”?
 - ❖ McNemar’s test (next slide)
 - ❖ Thurstonian model (subsequent slides)
- Experimental work needed to understand satisficing in sensory CATA (cf. Ares & Jaeger, 2013)



McNemar's Test

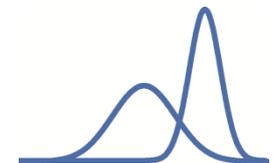
- Statistical analysis of CATA data in sensory not always correct
 - ❖ See Meyners, Castura, & Carr (2013) for review of correct analyses
- Common error when comparing products for a single attribute:
 - ❖ Treating evaluations as if they were independent

Product	Applies	Does Not Apply
A	70	30
B	80	20

- ❖ Should instead match data from each respondent

	Does Not Apply to B	Applies to B
Does Not Apply to A	10	20
Applies to A	10	60

- ❖ Analysis then proceeds using McNemar's test (or sign test)



Why is McNemar's Test Important?

- Full contingency tables distinguish between same marginal distributions

	Does Not Apply to B	Applies to B
Does Not Apply to A	0	30
Applies to A	20	50

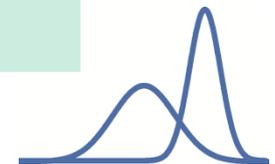
$p = 0.2$

	Does Not Apply to B	Applies to B
Does Not Apply to A	10	20
Applies to A	10	60

$p = 0.1$

	Does Not Apply to B	Applies to B
Does Not Apply to A	20	10
Applies to A	0	70

$p < 0.01$





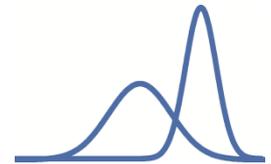
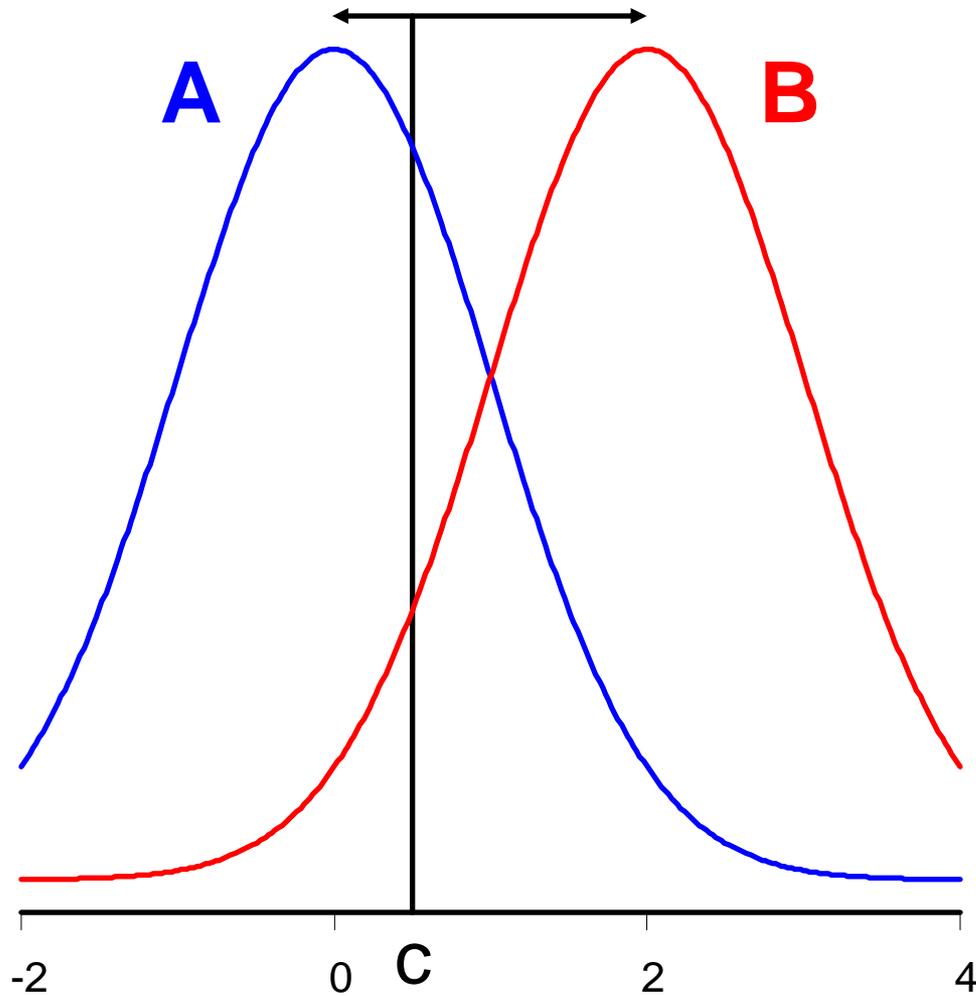
Scaling CATA Data

- Desirable to scale CATA data
 - ❖ If product A is selected twice as often as product B for an attribute, that does not mean that it has twice as much of that attribute
- Well-established scaling technique - Thurstonian modeling
 - ❖ Used to transform data to underlying interval scale
 - ❖ Allows comparisons across methods (Power, Precision, ...)
- Two main assumptions (Thurstone, 1927)
 - ❖ Perceptual noise
 - ❖ Decision rule
- Using Thurstonian modeling, we scale CATA data
 - ❖ We estimate the sensory difference (δ) between products on each attribute
 - ❖ We estimate the boundary (c) between “applies” and “does not apply”



Thurstonian model (δ and c)

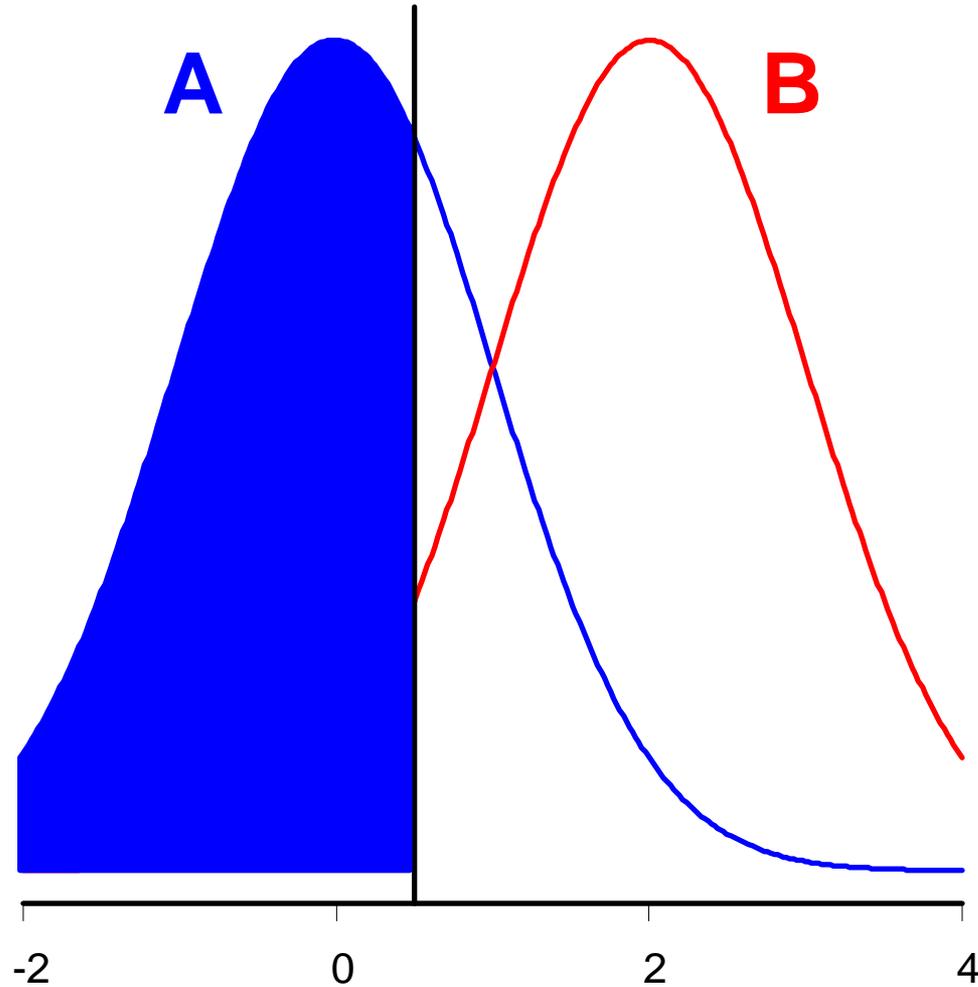
“Does Not Apply” ← | → δ “Applies”





“Does Not Apply – Product A”

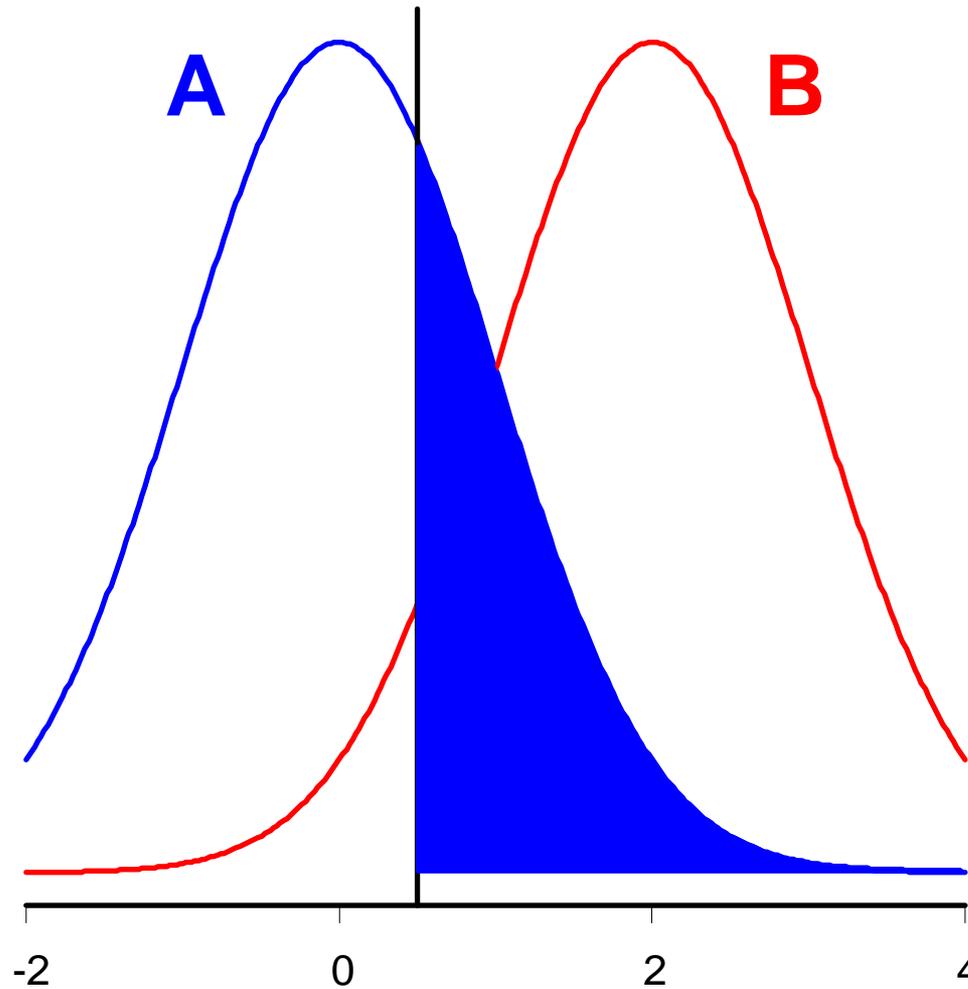
“Does Not Apply” ←|→ “Applies”





“Applies – Product A”

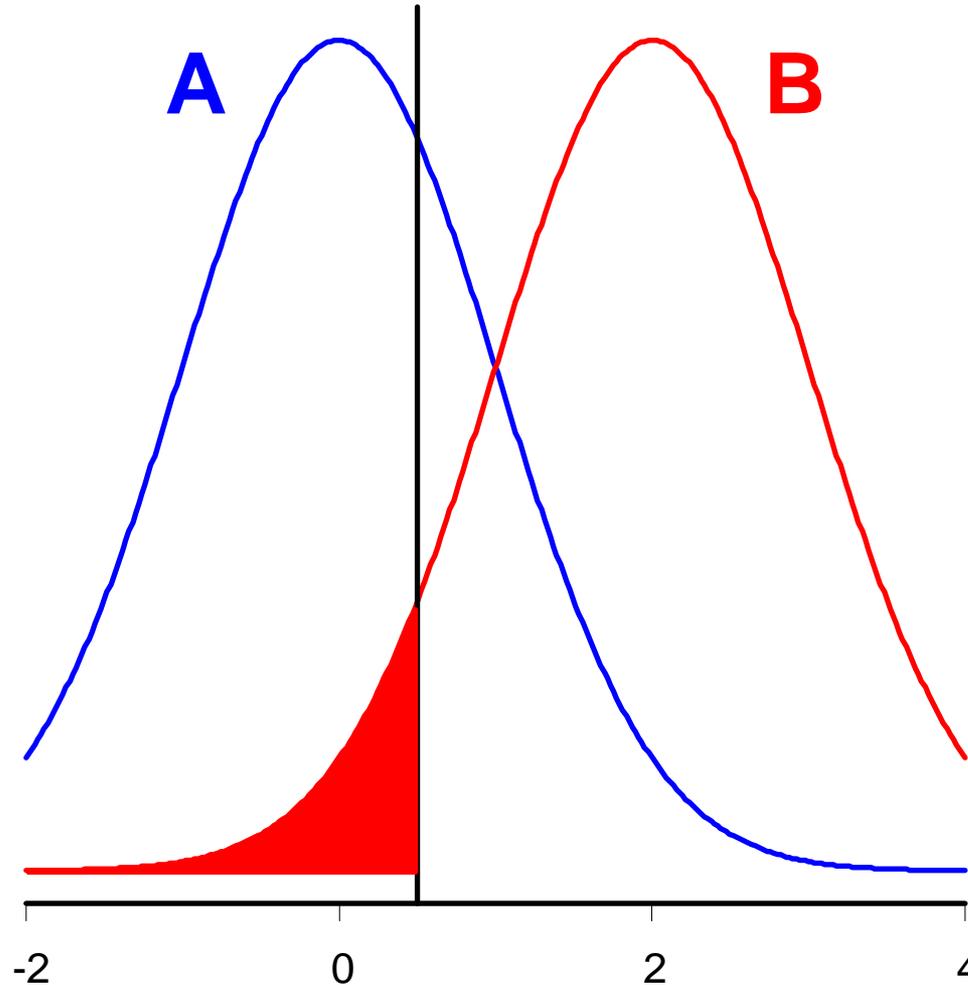
“Does Not Apply” ←|→ “Applies”





“Does Not Apply – Product B”

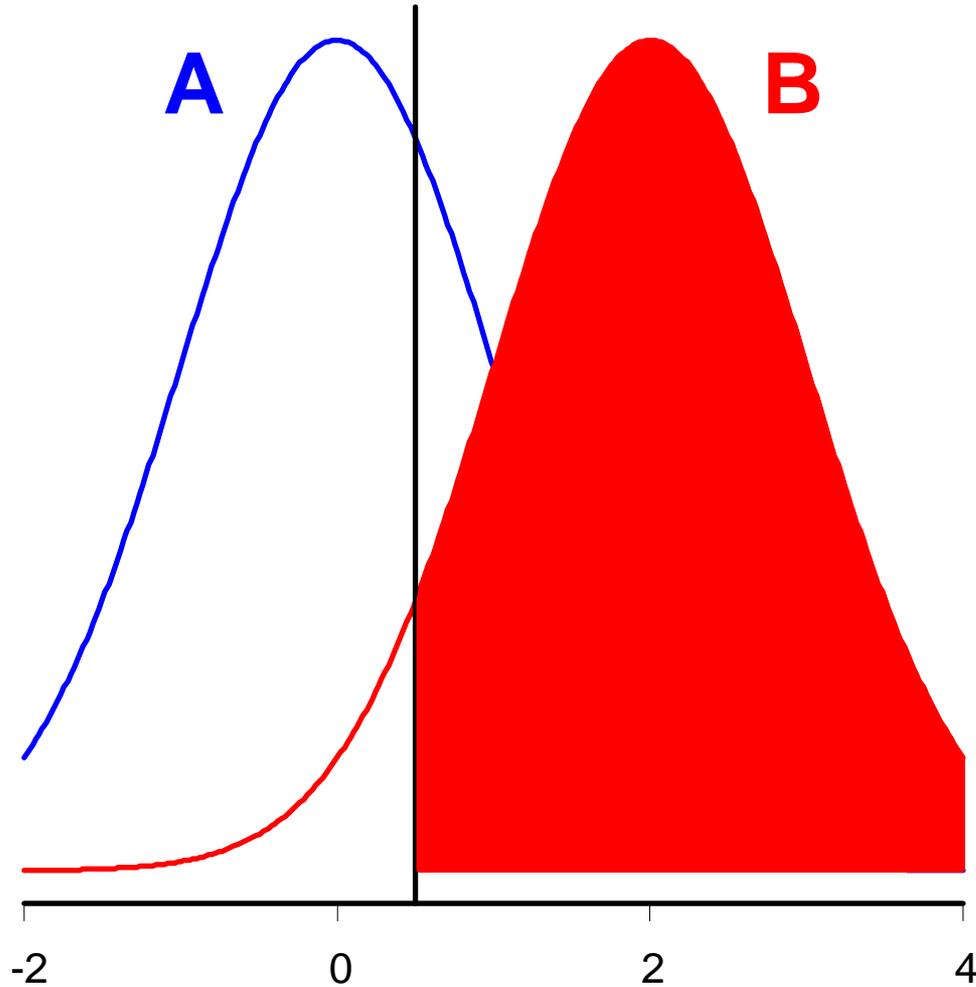
“Does Not Apply” \leftarrow | \rightarrow “Applies”





“Applies – Product B”

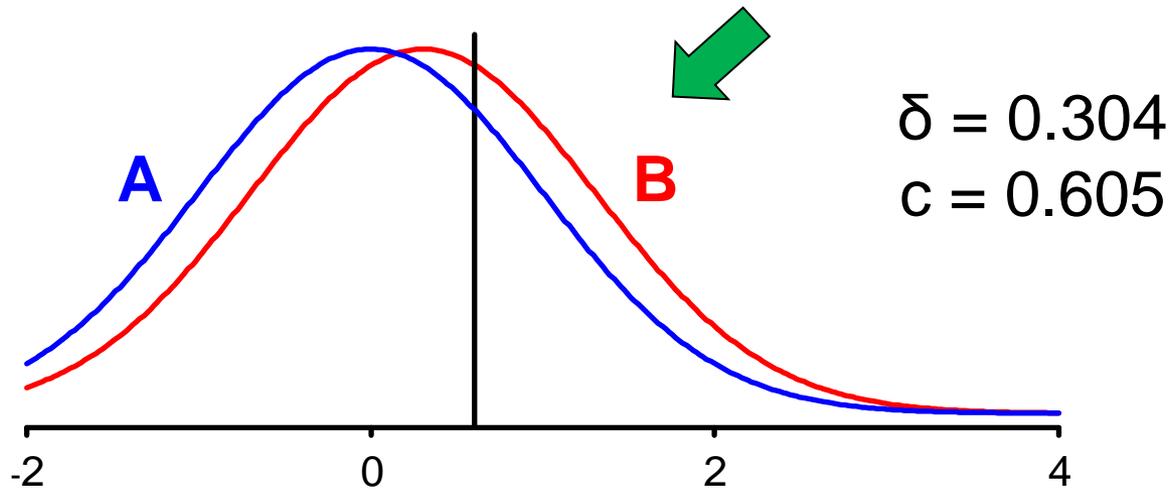
“Does Not Apply” ←|→ “Applies”



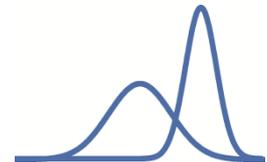
Example

- Fitting Thurstonian model to previous data

	Does Not Apply to B	Applies to B
Does Not Apply to A	10	20
Applies to A	10	60



- Issues:
 - ❖ Model assumes independent evaluations and same bound for all panelists
 - ❖ Same marginal distributions imply same estimates



Current State of Thurstonian Model

- Model currently implemented for two products on single attribute (Ennis and Ennis, 2013a)
- Model assumes same boundary and independent assessments for all respondents
 - ❖ Model for dependent evaluations in development
 - ❖ Extensions could consider different boundaries for different respondents (cf. Castura, Christensen, & Ennis, 2013)
- Extensions to three or more products needed
 - ❖ If no information desired regarding boundary, possible to use ranking model (Ennis and Ennis, 2013b)
- No model currently exists that spans multiple attributes

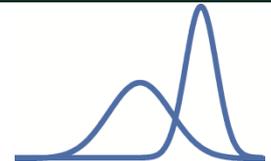


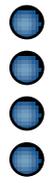


Future Directions and Summary

- Continued methodological investigations of CATA in sensory
 - ❖ Comparison of checklist and forced-choice CATA
- Greater awareness of statistical subtleties of CATA data
- Further development of Thurstonian model
 - ❖ Extensions to model
 - ❖ Panel tracking
 - ❖ Power investigations
 - ❖ Precision of measurement
 - ❖ Models to span multiple attributes

Summary: CATA is an efficient technique for collecting meaningful sensory data but more work is needed

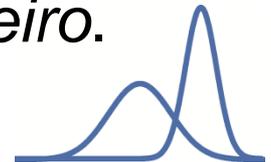




Resources and References (1/4)

This talk will be available on www.ifpress.com, and the March 2014 IFP webinar will discuss this topic in greater detail.

- Ares, G., & Jaeger, S. R. (2013). Check-all-that-apply questions: Influence of attribute order on sensory product characterization. *FQ&P*, 28(1), 141–153.
- Ares, G., Jaeger, S. R., Bava, C. M., Chheang, S. L., Jin, D., Gimenez, A., ... Varela, P. (2013). CATA questions for sensory product characterization: raising awareness of biases. *FQ&P*.
- Castura, J. C., Ennis, J. M., & Christensen, R. H. B. (2013). Panel Tracking via Thurstonian Modeling. *Poster Presented at 10th Annual Pangborn Symposium, Rio de Janeiro.*



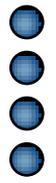


Resources and References (2/4)

This talk will be available on www.ifpress.com, and the March 2014 IFP webinar will discuss this topic in greater detail.

- Ennis, D., & Ennis, J. (2013). A Thurstonian ranking model with rank-induced dependencies. *Journal of Classification*, 30(1), 124–147.
- Ennis, D. M., & Ennis, J. M. (2013). Analysis and Thurstonian scaling of applicability scores. *JSS*, 28(3), 188–193.
- Jaeger, S. R., Chheang, S. L., Jin, D., Bava, C. M., Gimenez, A., Vidal, L., & Ares, G. (2013). Check-all-that-apply (CATA) responses elicited by consumers: within-assessor reproducibility and stability of sensory product characterizations. *FQ&P*.

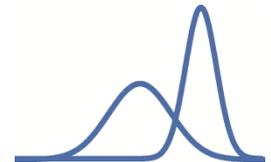


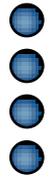


Resources and References (3/4)

This talk will be available on www.ifpress.com, and the March 2014 IFP webinar will discuss this topic in greater detail.

- Loh, K., & Ennis, D. M. (1982). Glossary of Attributes. *Philip Morris Technical Report*, 82(288). Retrieved from http://www.pmdocs.com/pdf/2001298236_8263_120_201326123634.pdf
- Krosnick, J. A. (1999). Survey research. *Annual review of psychology*, 50(1), 537–567.
- Meyners, M., Castura, J. C., & Thomas Carr, B. (2013). Existing and new approaches for the analysis of CATA data. *FQ&P*.





Resources and References (4/4)

This talk will be available on www.ifpress.com, and the March 2014 IFP webinar will discuss this topic in greater detail.

- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, 70(1), 66–77.
- Sudman, S., & Bradburn, N. M. (2012). Asking questions: A practical guide to questionnaire design.
- Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286.





Acknowledgments

- **Daniel Ennis**, The Institute for Perception, USA
- **Michael Meyners**, Proctor & Gamble, Germany
- **Gaston Ares**, Universidad de la Republica, Uruguay
- **Sara Jaeger**, The New Zealand Institute for Plant & Food Research, New Zealand
- **You!**





**10th Pangborn Sensory
Science Symposium**
Rio de Janeiro, Brazil
August 15th, 2013



***Developments in Applicability and
CATA scoring***

John M. Ennis

Daniel M. Ennis

The Institute for Perception

john.m.ennis@ifpress.com

daniel.m.ennis@ifpress.com

