

**Sensory Hub Conference**  
**February 9, 2026**



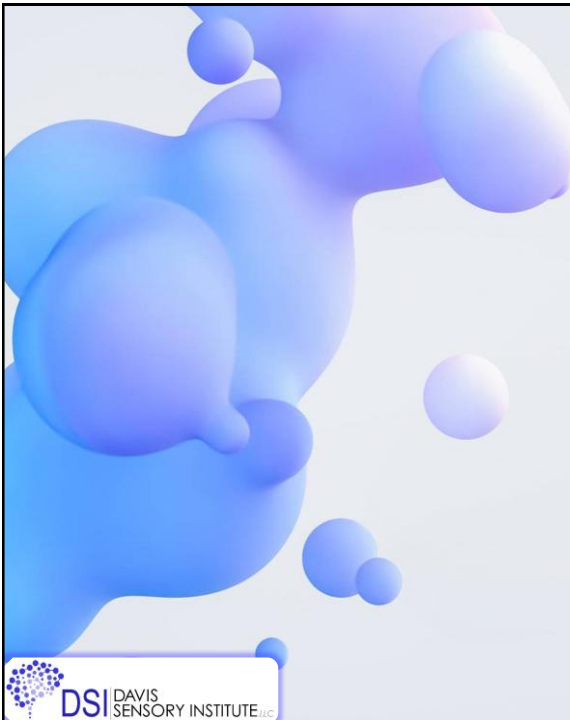
# **Closing the Loop: Integrating Trained Panels with Consumer Perception**

*Benoît Rousseau, Ph.D.*

*President*


*The Davis Sensory Institute*








Closing the Loop:  
Integrating  
Trained Panels  
with  
Consumer Perception

# Illustrating Scenario



2/22


2


- Major orange juice manufacturer
- Reformulation of their main high-pulp OJ brand due to weather-related crop issues in Florida
- Must increase ratio of California oranges
- Research will involve 3 prototypes with increasing California fruit content

Product	CA/FL Ratio
Control	20/80
Prototype 1	25/75
Prototype 2	30/70
Prototype 3	35/65

**Objective**  
Identify the largest ratio not resulting in a meaningful sensory difference from control



- Two-step research approach
  - **Internal:** Tetrad method with trained panelists
  - **Consumer:** Confirmatory preference tests



3/22

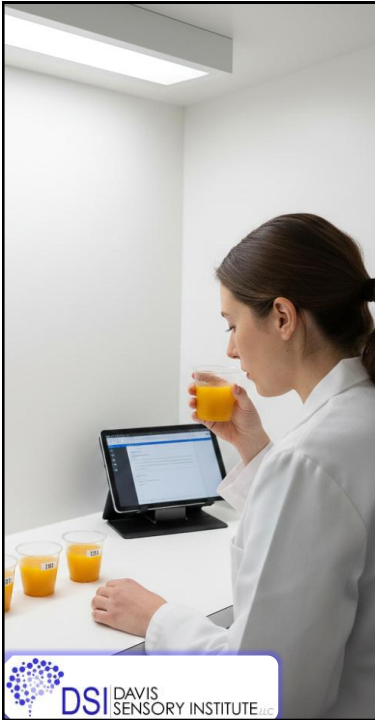



3

## Internal Tetrad Testing

- 24 highly trained discrimination testing panelists
- One tetrad per panelist for each comparison
- Binomial,  $p_0 = 33\%$ 
  - Number for significance at the 0.05 level: Need at least  $13/24$

Comparison		# correct	p-value
Control	Prototype 1 <i>(25/75)</i>	9 / 24	> 0.05
	Prototype 2 <i>(30/70)</i>	11 / 24	> 0.05
	Prototype 3 <i>(35/65)</i>	14 / 24	< 0.05

**Conclusion:**  
 Up to a CA/FL 30/70 ratio is acceptable

4/22

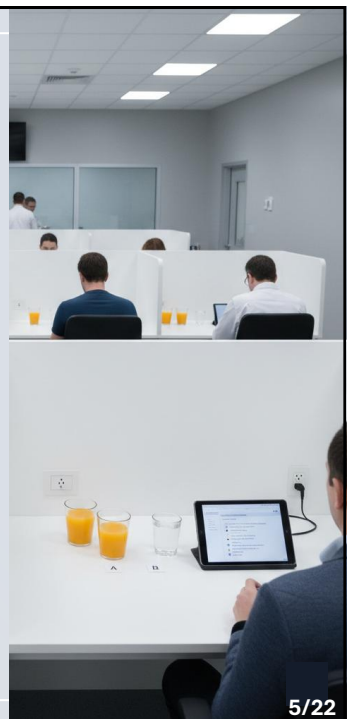



4

## Consumer Testing

- 80 regular users of the company's high-pulp OJ brand
- One paired preference per consumer (forced choice) of each comparison
- Binomial,  $p_0 = 50\%$ 
  - Number for significance at the 0.05 level (one-tailed): Need at least  $48/80$
- Results


Comparison		Control chosen	p-value
Control	Prototype 1 <i>(25/75)</i>	43 / 80 <i>(54%)</i>	> 0.05
	Prototype 2 <i>(30/70)</i>	51 / 80 <i>(64%)</i>	< 0.01
	Prototype 3 <i>(35/65)</i>	59 / 80 <i>(74%)</i>	< 0.001

**Conclusion:**  
 Up to a CA/FL 25/75 ratio is acceptable









5/22

5



# Decision Conundrum

TETRAD				PREFERENCE	
Comparison	# correct	p-value		p-value	# Control
Prototype 1 <i>(25/75)</i>	9 / 24	> 0.05		> 0.05	43 / 80 <i>(54%)</i>
Control Prototype 2 <i>(30/70)</i>	11 / 24	> 0.05		< 0.001	51 / 80 <i>(64%)</i>
Prototype 3 <i>(35/65)</i>	14 / 24	< 0.05		< 0.001	59 / 80 <i>(74%)</i>

How can we explain the **Prototype 2** results?





- Are the internal panelists less sensitive than consumers?
- Is discrimination testing not able to predict consumer preferences?

1. SAMPLE SIZE

3. CONSUMER RELEVANCE

2. METHODOLOGY

4. TRAINED VS. CONSUMER

6/22

6

1. SAMPLE SIZE

2. METHODOLOGY

3. CONSUMER RELEVANCE

4. TRAINED VS. CONSUMER



**Closing the Loop:  
Integrating  
Trained Panels  
with  
Consumer Perception**



# SAMPLE SIZE







7/22

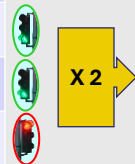
7

# Sample Size

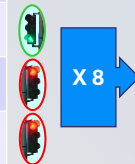
- Two samples will **never** be identical
- **Chew (1977)**: Two putatively different samples will always be found to be statistically different given a sample size that is large enough
- Case in point (OJ tetrad):



Control vs.	# correct	p-value
Prot. 1	9 / 24 (38%)	> 0.05
Prot. 2	11 / 24 (46%)	> 0.05
Prot. 3	14 / 24 (58%)	< 0.05



Control vs.	# correct	p-value
Prot. 1	18 / 48 (38%)	> 0.05
Prot. 2	22 / 48 (46%)	< 0.05
Prot. 3	28 / 48 (58%)	< 0.05



Control vs.	# correct	p-value
Prot. 1	146 / 384 (38%)	< 0.05
Prot. 2	176 / 384 (46%)	< 0.05
Prot. 3	224 / 384 (58%)	< 0.05

# Decision Conundrum



TETRAD				PREFERENCE	
Comparison	# correct	p-value		p-value	# Control
Prototype 1 (25/75)	9 / 24	> 0.05		> 0.05	43 / 80 (54%)
Control Prototype 2 (30/70)	11 / 24	> 0.05		< 0.001	51 / 80 (64%)
Prototype 3 (35/65)	14 / 24	< 0.05		< 0.001	59 / 80 (74%)



1. SAMPLE SIZE  
2. METHODOLOGY  
3. CONSUMER RELEVANCE  
4. TRAINED VS. CONSUMER

Closing the Loop: Integrating Trained Panels with Consumer Perception

# METHODOLOGY

TRIANGLE TETRAD 2-AFC

DSI DAVIS SENSORY INSTITUTE LLC 10/22

## Methodology

- Methodologies differ in their ability to detect sensory differences
- Driven by
  - Inner-sample variability
  - Specific task decision rule

Thurstonian Framework

Delta (d')	2-AFC Power (%)	Tetrad Power (%)	Triangle Power (%)	Duo-trio Power (%)
0.0	0	0	0	0
0.5	30	10	5	2
1.0	80	50	25	10
1.5	95	90	70	40
2.0	100	100	95	80

Power

Delta (d'): Size of sensory difference

2-AFC Tetrad Triangle Duo-trio

DSI DAVIS SENSORY INSTITUTE LLC 11/22

# Methodology

**Thurstonian Framework**

• Expected outcomes from OJ research:

Control vs.	# correct	p-value
Prot. 1	9 / 24	> 0.05
Prot. 2	11 / 24	> 0.05
Prot. 3	14 / 24	< 0.05

Control vs.	# correct	p-value
Prot. 1	8 / 24	> 0.05
Prot. 2	10 / 24	> 0.05
Prot. 3	11 / 24	> 0.05

Control vs.	# correct	p-value
Prot. 1	15 / 24	> 0.05
Prot. 2	18 / 24	< 0.05
Prot. 3	20 / 24	< 0.05



12/22

1. SAMPLE SIZE
2. METHODOLOGY
3. CONSUMER RELEVANCE
4. TRAINED VS. CONSUMER


**Closing the Loop:  
Integrating  
Trained Panels  
with  
Consumer Perception**

# CONSUMER RELEVANCE

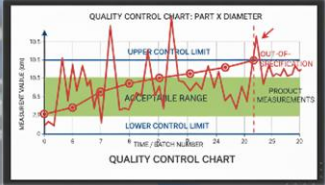
13/22





# Fundamental Principle




- Two samples will **never** be identical
- Instead of checking for ‘matching’ (impossible), we need to establish ‘small enough’
  - Concept identical to that broadly used in **quality control**





14/22


## How Can We Establish the Upper Sensory Control Limit?

Upper Sensory Control Limit




No difference





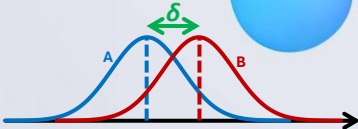
Preference Test


- How big of a difference before consumers have a preference?





Same-Different test

- How big of a difference before consumers call two products “different”?



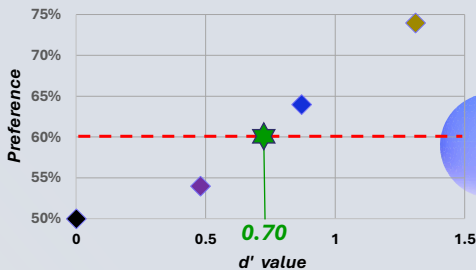
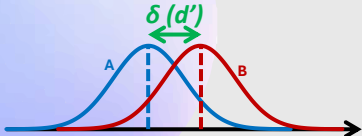



15/22




## Illustration with Scenario

		TETRAD		PREFERENCE	
Comparison		# correct	d' value	# Control	
<b>Control</b>	<b>Prototype 1</b>	<b>9 / 24</b>	<b>0.48</b>	43 / 80 <i>(54%)</i>	
	<b>Prototype 2</b>	<b>11 / 24</b>	<b>0.87</b>	51 / 80 <i>(64%)</i>	
	<b>Prototype 3</b>	<b>14 / 24</b>	<b>1.31</b>	59 / 80 <i>(74%)</i>	

- **Notes on using preference to establish consumer relevance**
  - Need to set the preference 'threshold' (statistical significance not sufficient since contingent on sample size)
  - Sensory differences can trigger population segmentation and thus lower estimate of the underlying preference
- **We generally recommend using the same-different method**



16/22

16

1. SAMPLE SIZE

2. METHODOLOGY

3. CONSUMER RELEVANCE

4. TRAINED VS. CONSUMER



# Building the Consumer – Internal Relationship



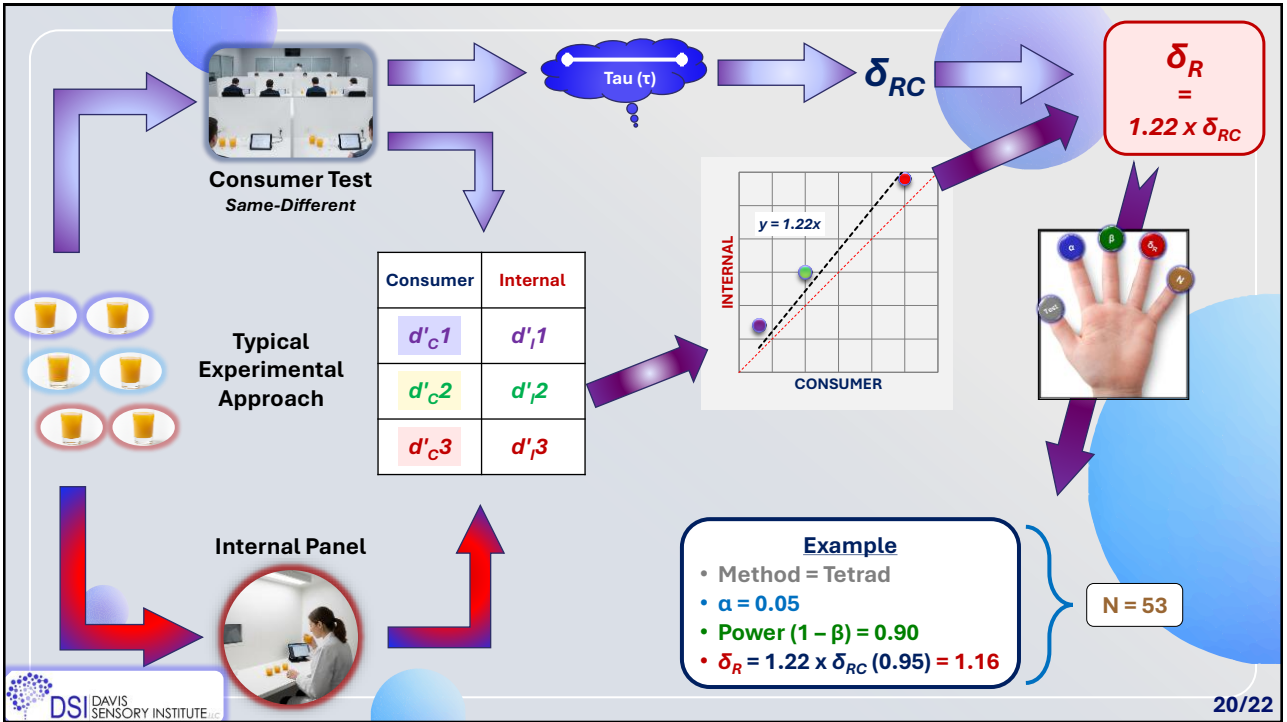


17/22

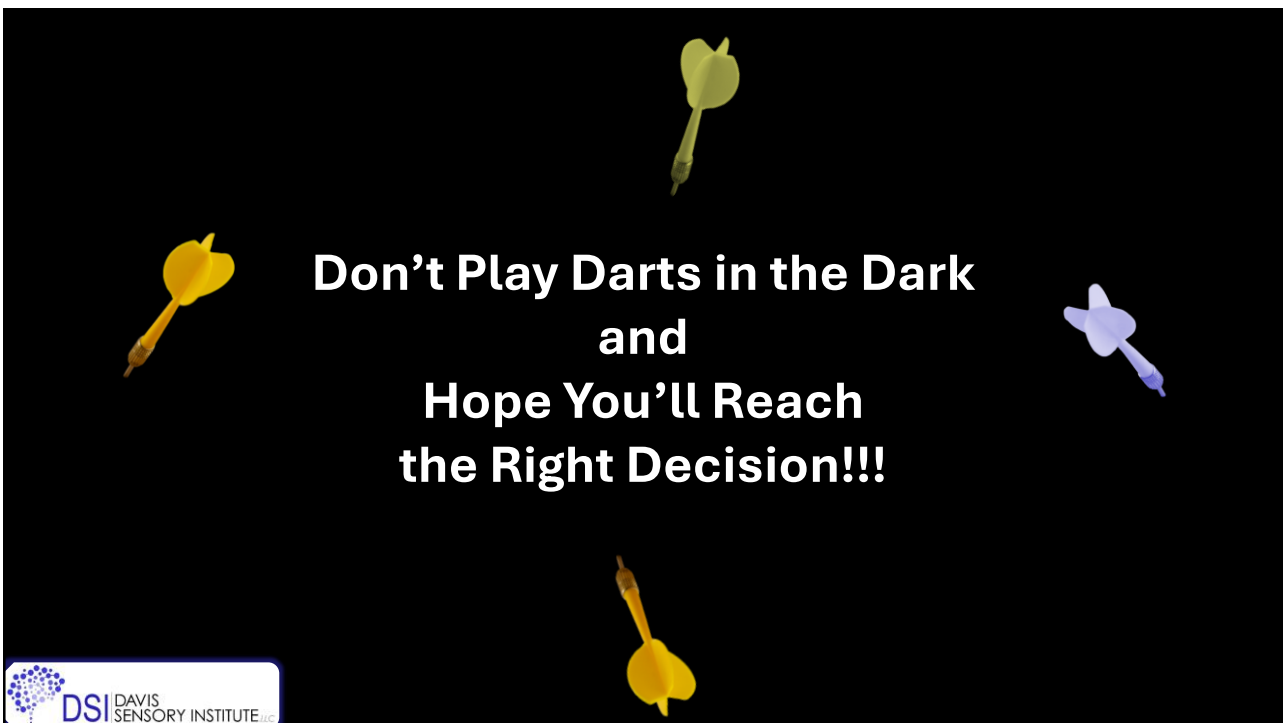
17

8





20



21



Sensory Hub Conference  
February 9, 2026



# Closing the Loop: Integrating Trained Panels with Consumer Perception

*Benoît Rousseau, Ph.D.*

*President*

*The Davis Sensory Institute*



22/22