

# CONSUMER DIVERSITY IN SENSORY EVALUATION DATA

**John C. Castura**  
VP Innovation & Research

AGROSTAT 2018, édition 2018



Decisions related to experimental design and statistical analysis in sensory evaluation are often guided by standard practices.

Does a study that is designed and analyzed in a manner that is consistent with these practices always make sense?



**“A foolish consistency is  
the hobgoblin of little  
minds...”**

**Ralph Waldo Emerson  
1841**





Who is an  
“average consumer”?





# **Part I:**

## **Hedonic data**



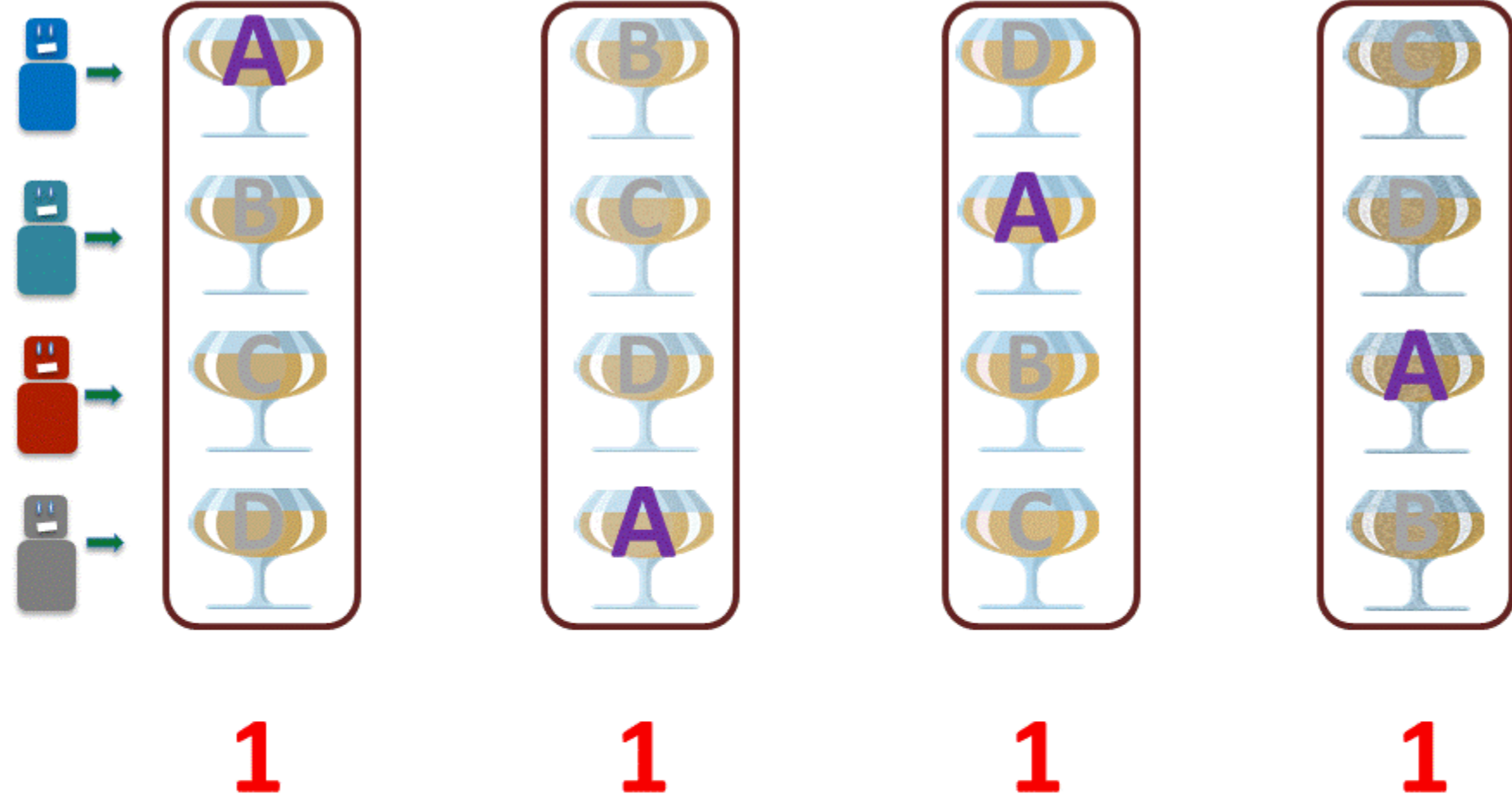


# Part I: Consumer acceptance

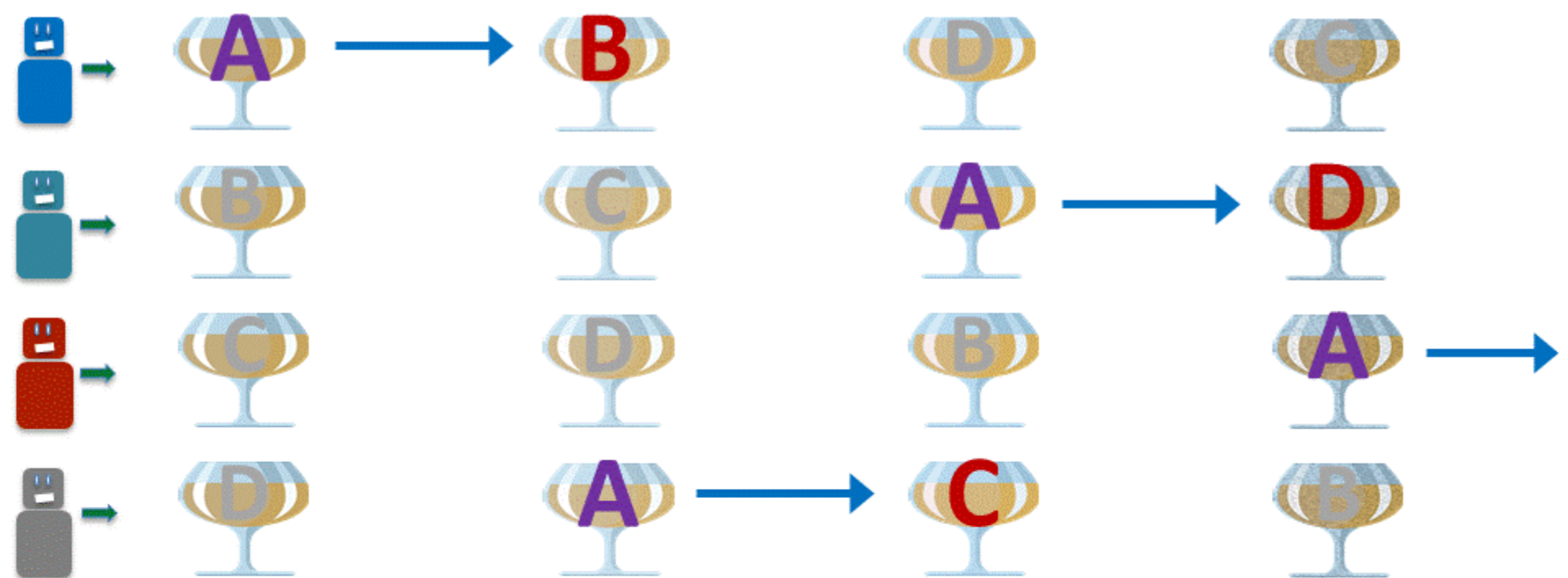


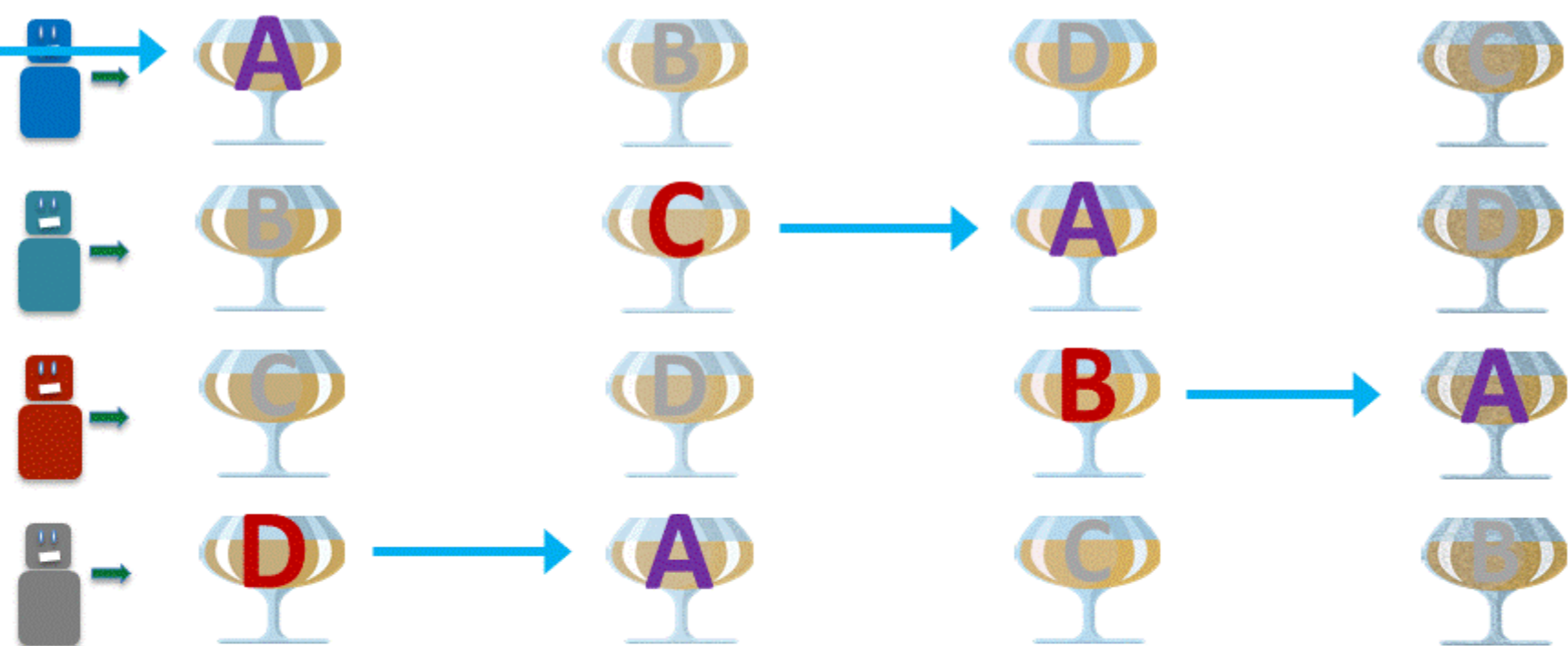




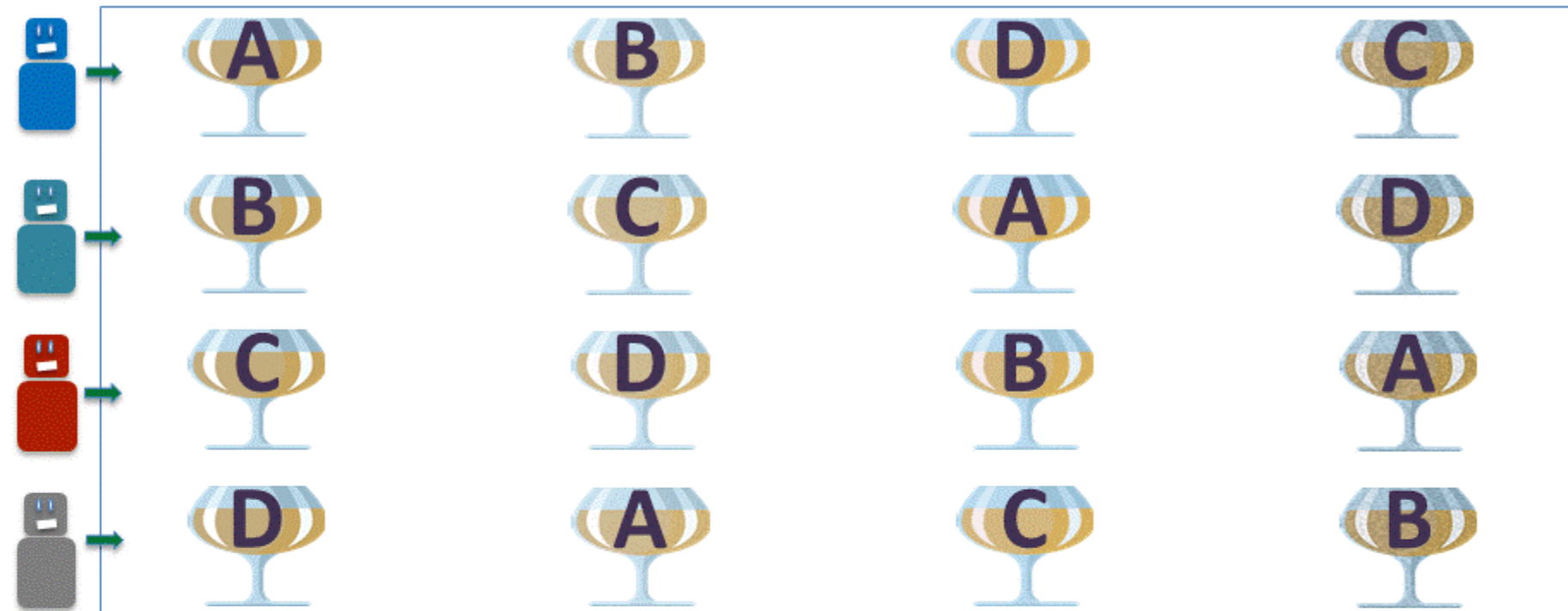




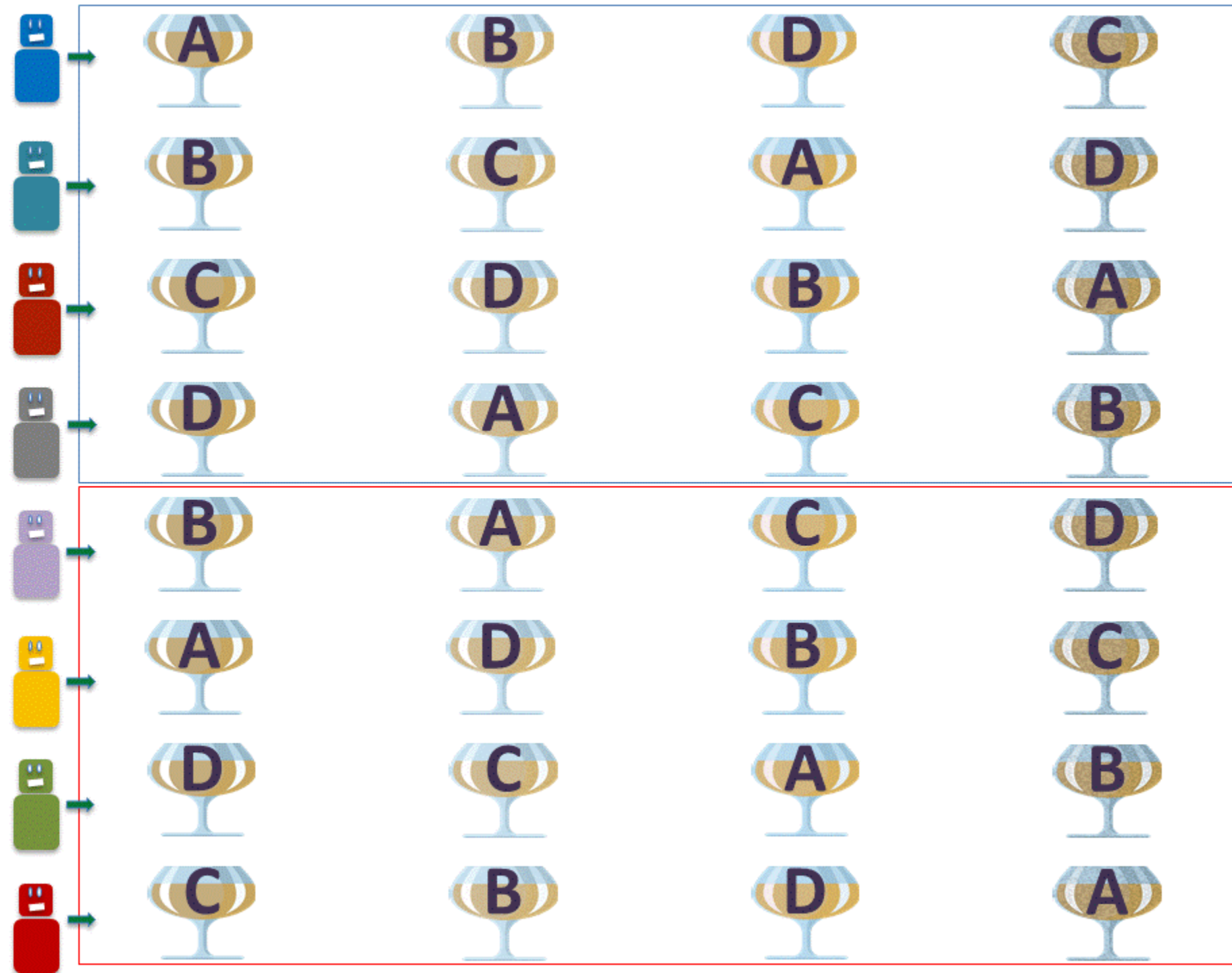






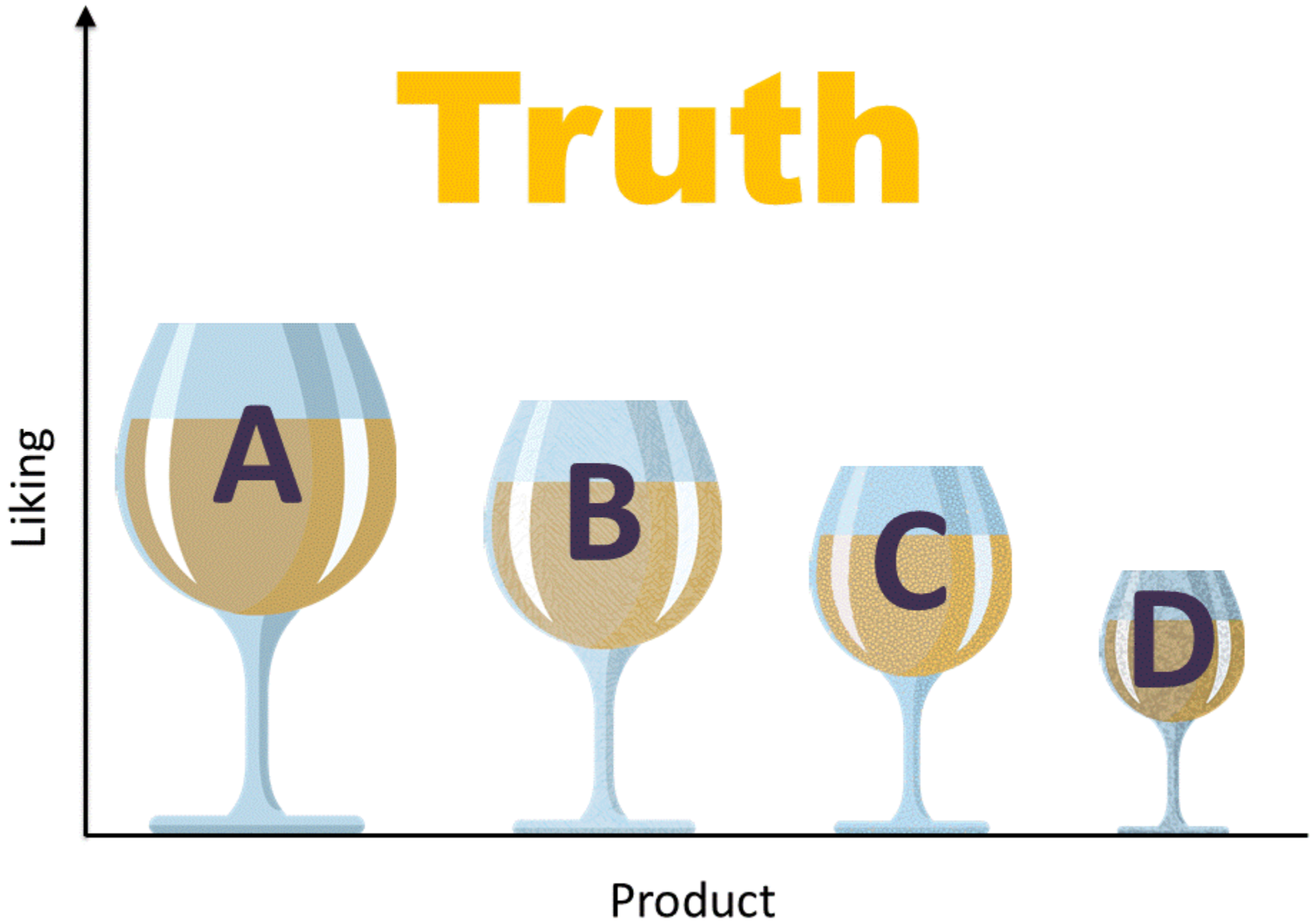


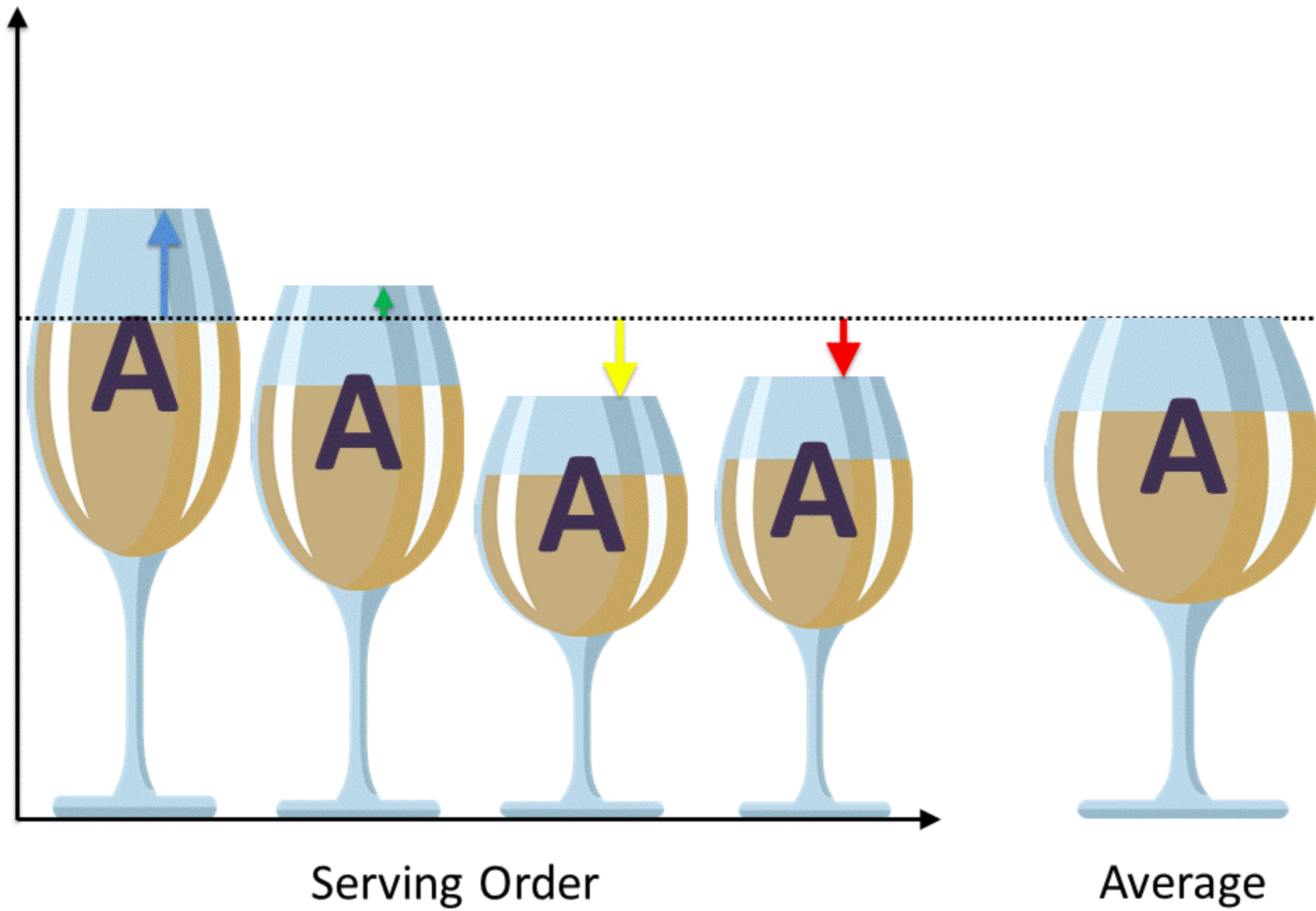
Williams design (4 treatments)



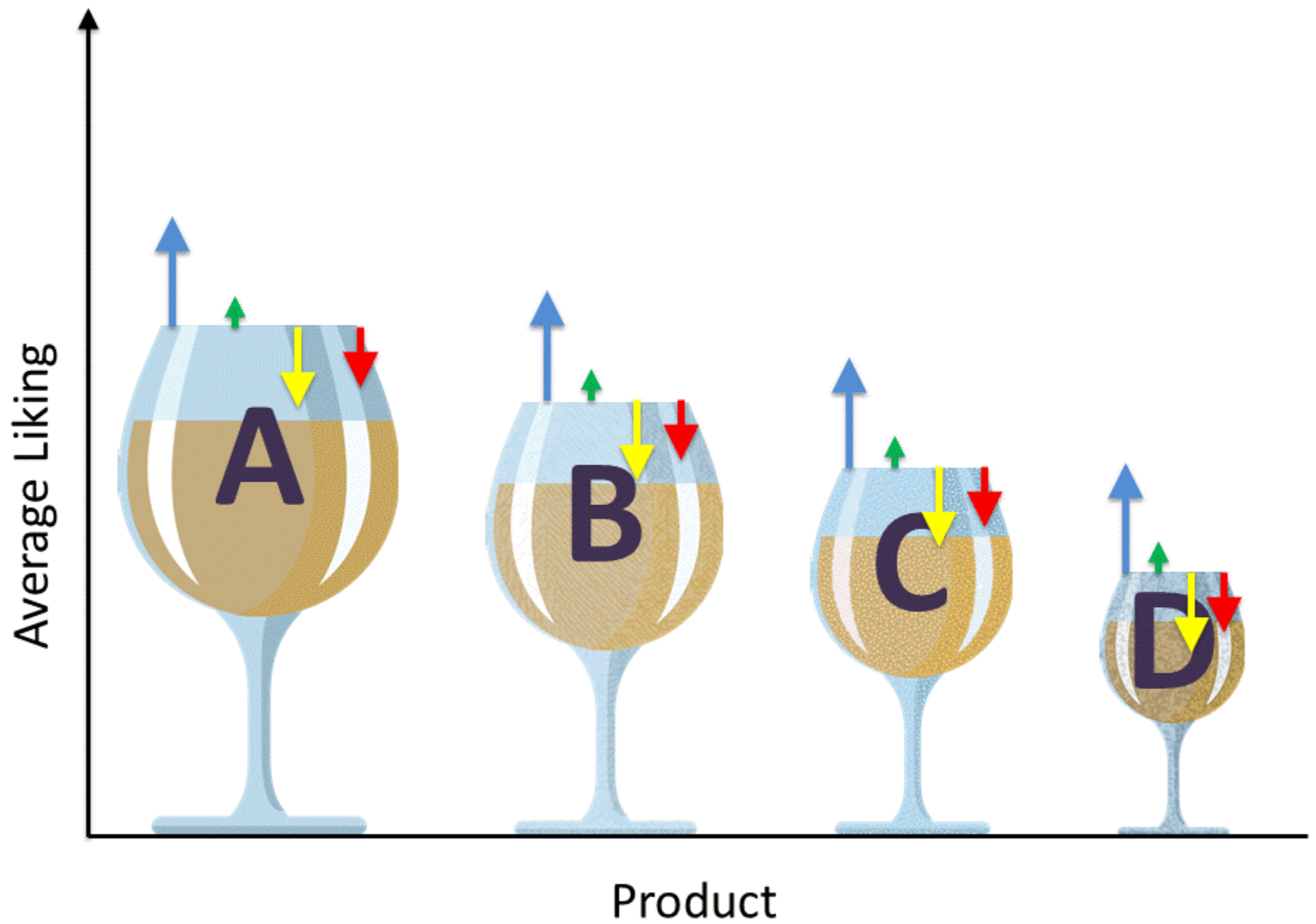


# Truth









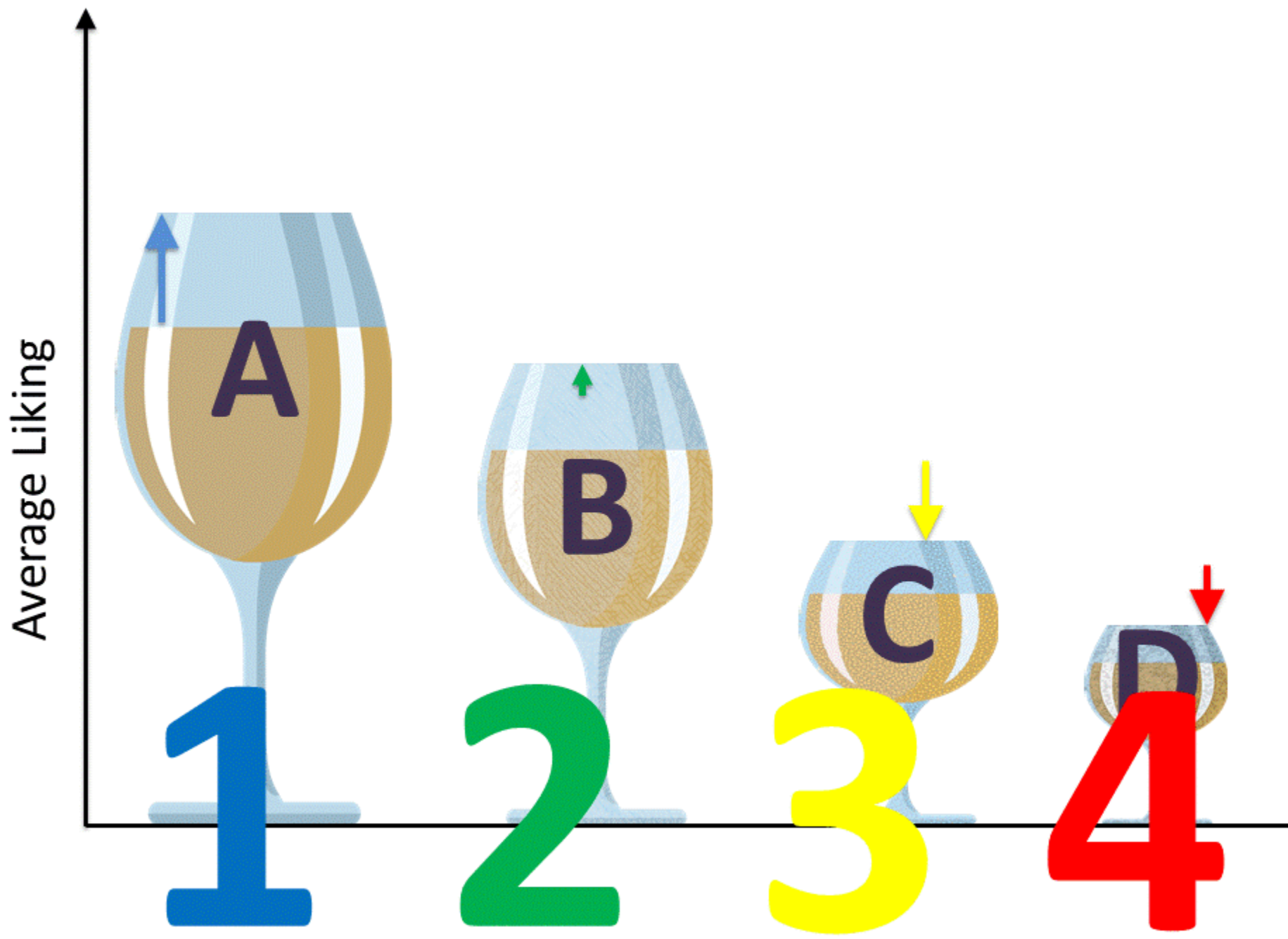


1

2

3





Average Liking







Liking

1

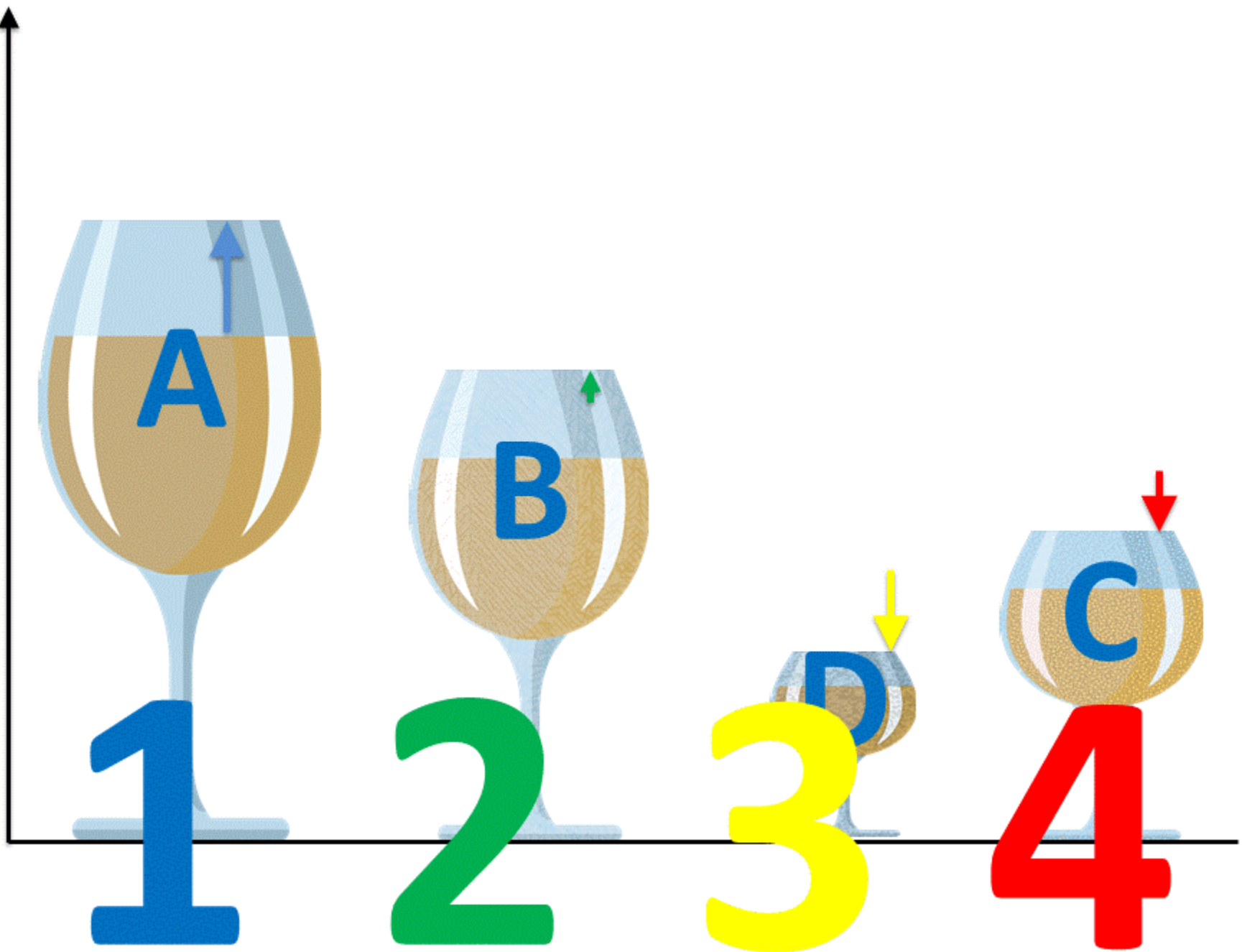
2

3

4

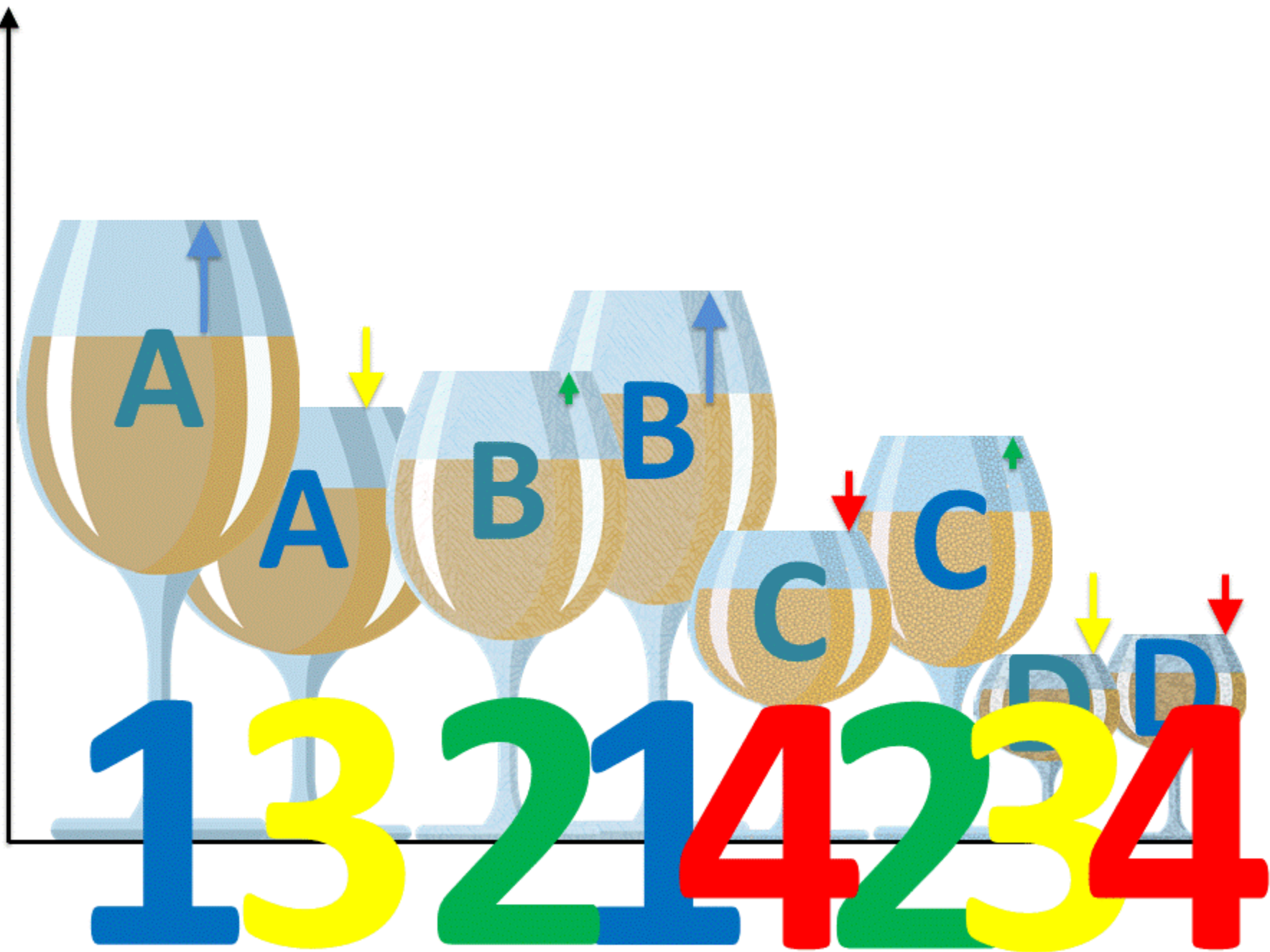


Liking

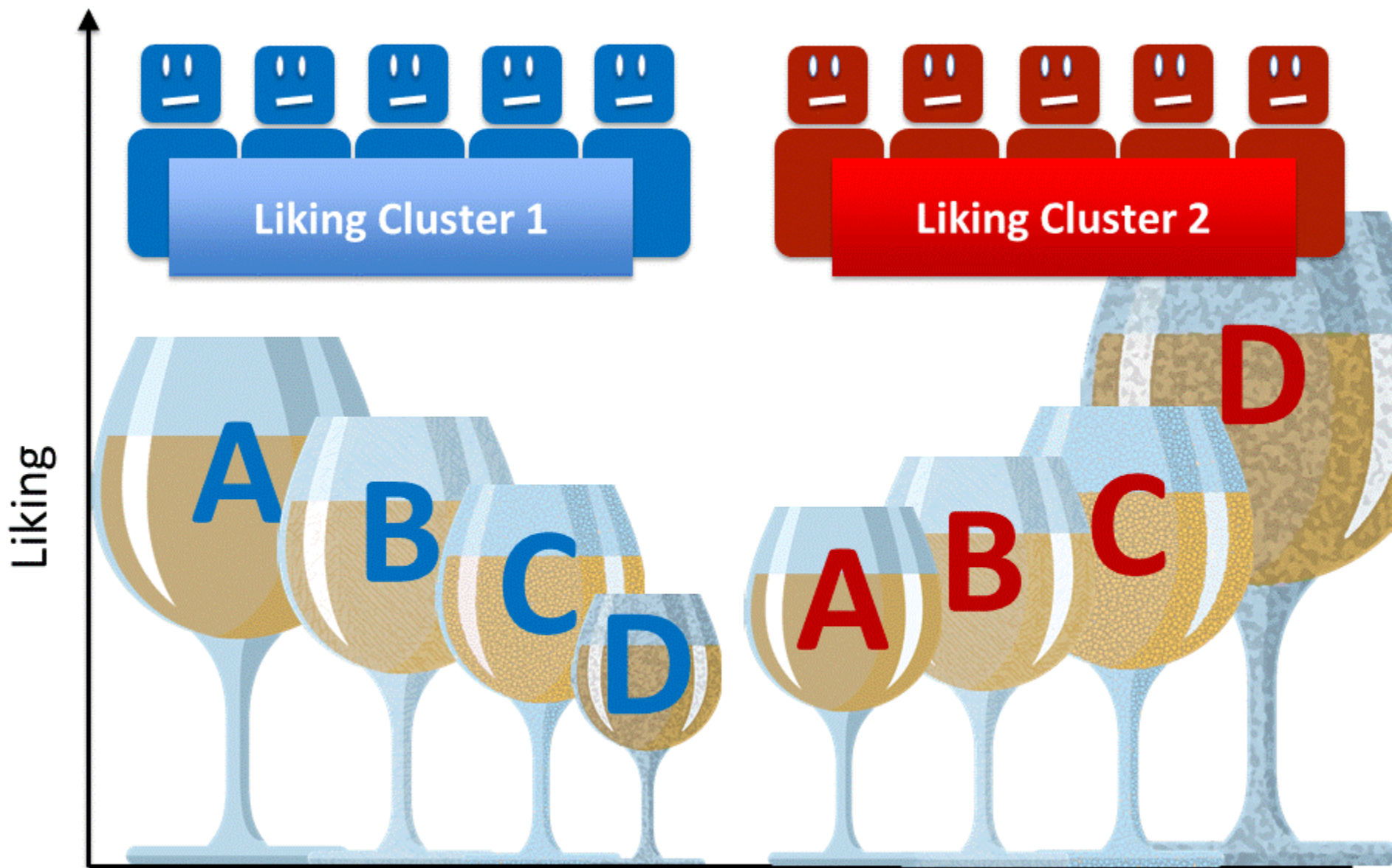




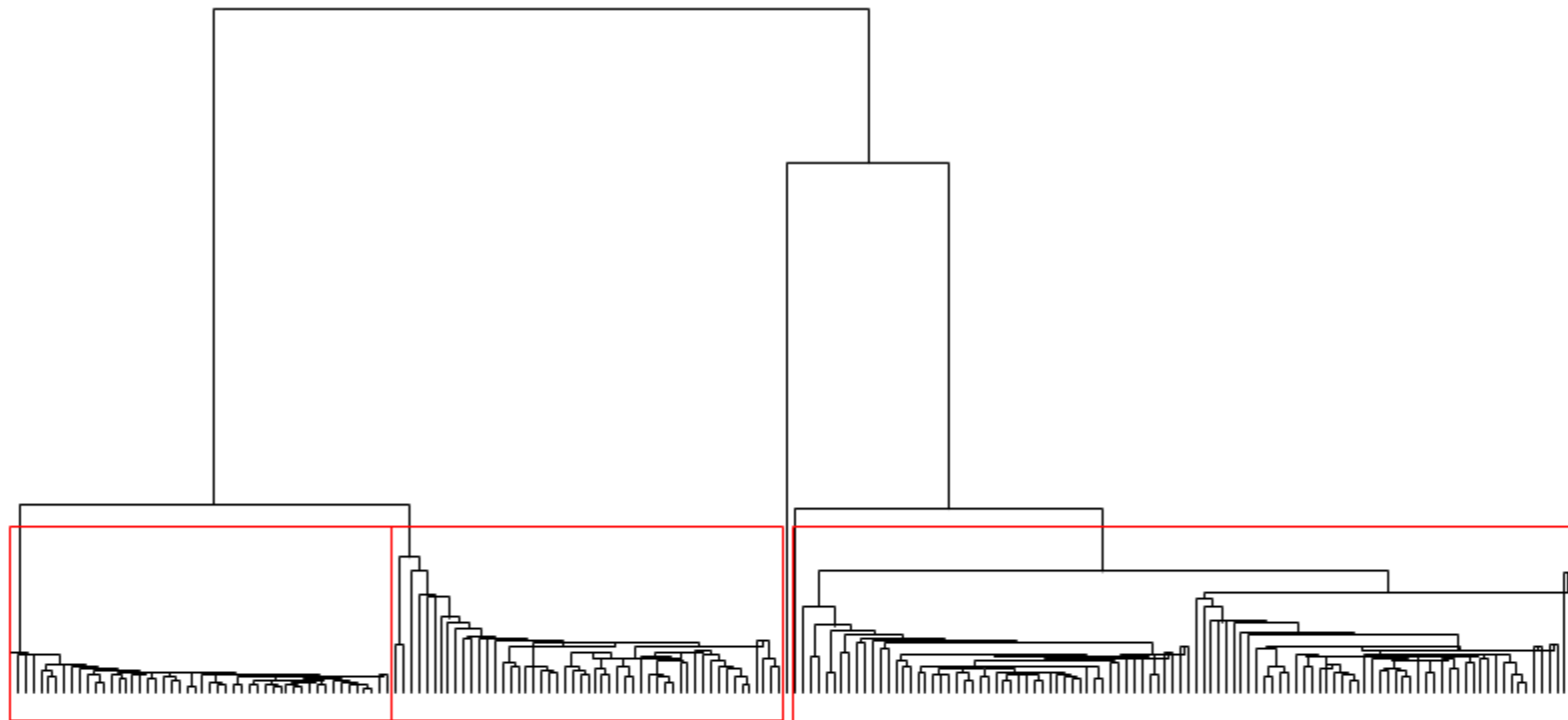
Liking



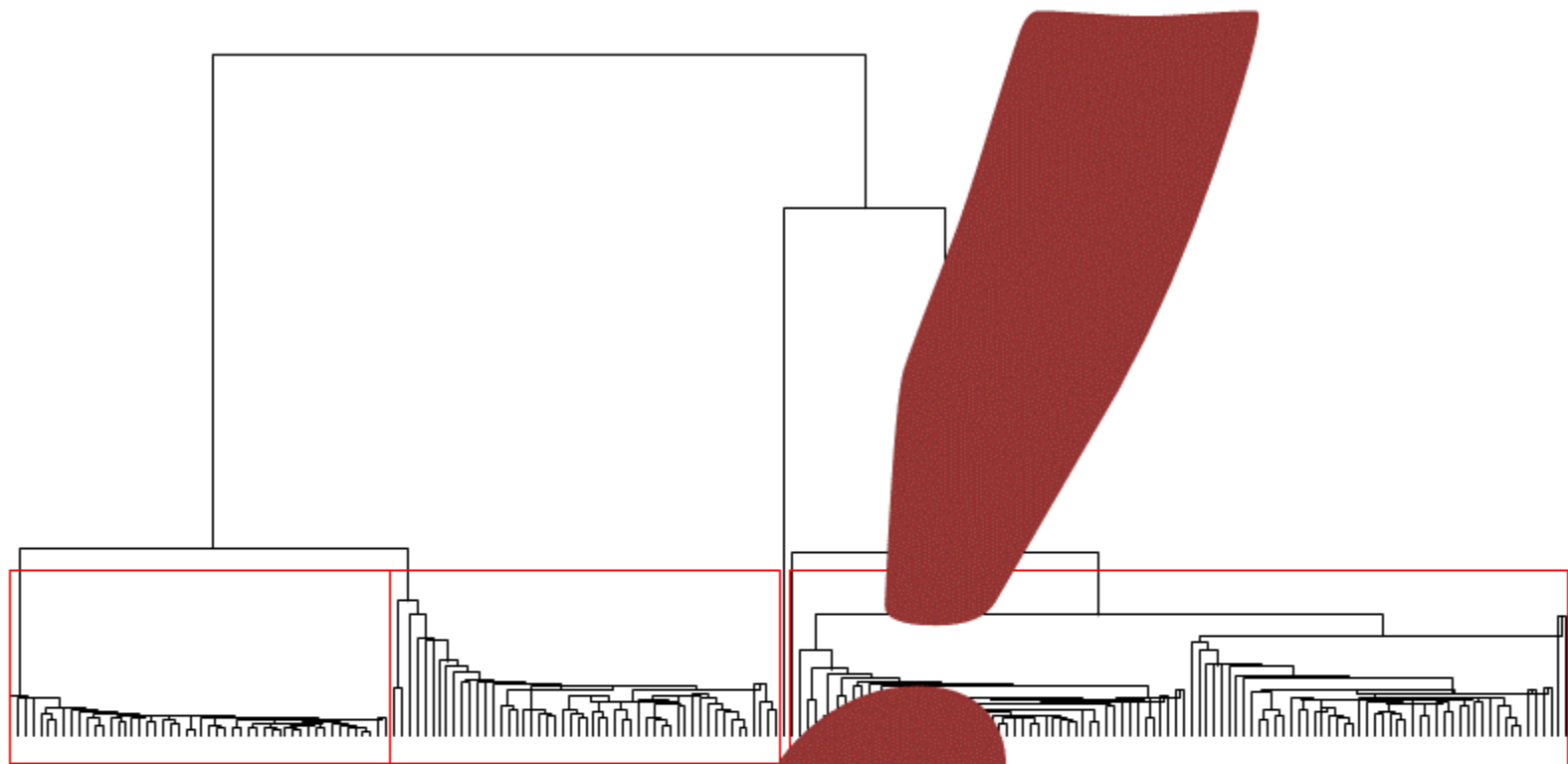




# HCA



HCA



First  
sample  
was...

**A**

**D**

**B or C**



Various authors have reported situations in which consumers are clustered according to the randomly allocated serving orders.

**Think about that when using this data in preference mapping!**

Liking responses will be influenced by context effects and various biases. Thus we should think of liking responses as ***momentary*** and not as a fixed property of the consumer.

Also **replication** of consumer panel results seems more important than **repeatability** of individual consumers.

# Liking data

organize (scaled?) liking data into a 2D array

**Rows: Consumers**

**Columns: Products**

×

A

B

C

D

5

4

5

5

8

5

6

6

7

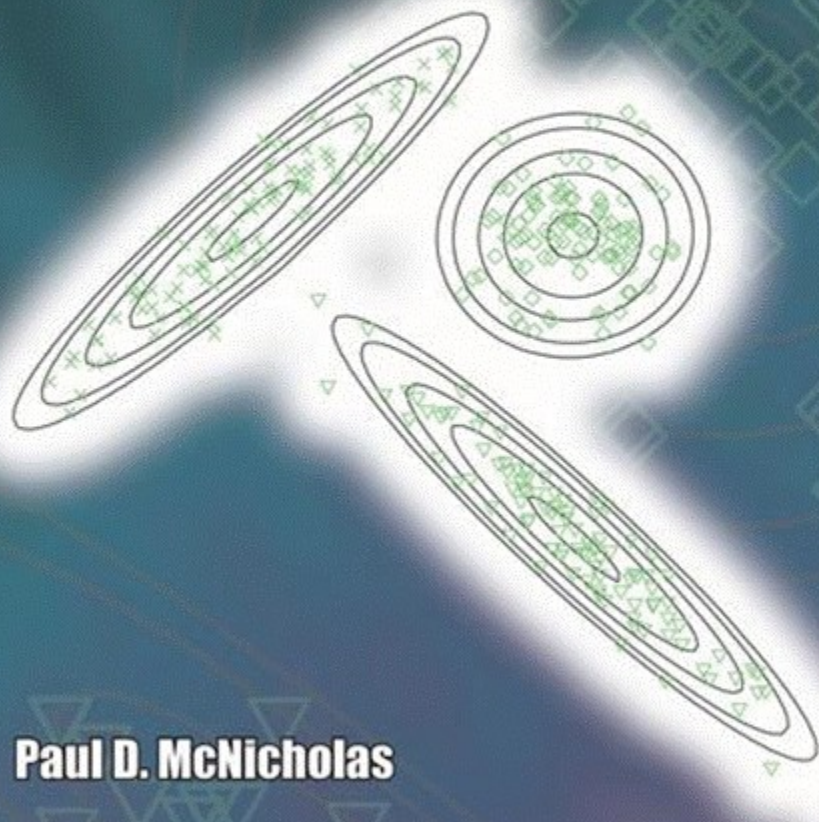
6

6

7



# MIXTURE MODEL-BASED CLASSIFICATION



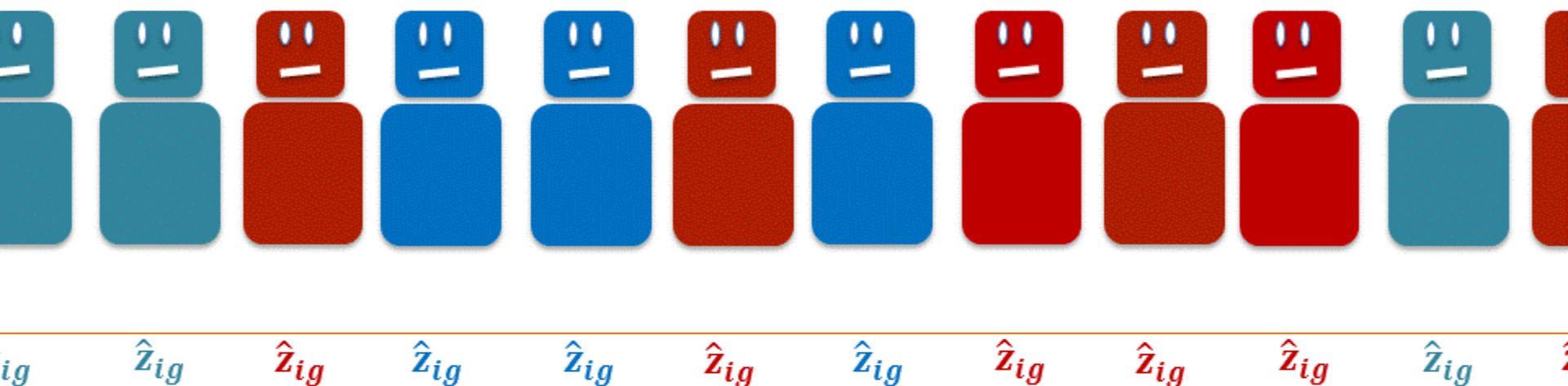
**Paul D. McNicholas**



CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

# Gaussian Mixture Model



Initialize  $\hat{z}_{ig}$ .

**M step** – update  $\hat{\pi}_g, \hat{\mu}_g, \hat{\Sigma}_g$ .

**E step** – update  $\hat{z}_{ig}$  classification predictions.  
(Stop when converged.)

# Higher dimensional data

Attempt to relate

**observed variables ( $p$ )**

to

**latent variables ( $q$ )**

where  $q < p$  ...and perhaps  $q \ll p$ .



# Mixture of Factor Analyzers

$$\mu_g, \Sigma_g = \Lambda_g \Lambda_g' + \Psi_g$$

**Group 1:**  $\mu, \Sigma = \Lambda \Lambda' + \Psi$

**Group 2:**  $\mu, \Sigma = \Lambda \Lambda' + \Psi$

**Group 3:**  $\mu, \Sigma = \Lambda \Lambda' + \Psi$

# Parsimonious Gaussian Mixture Model

$$\mu_g, \Sigma_g = \Lambda_g \Lambda_g' + \omega_g \Delta_g$$

with CUU constraints

$$\Lambda_g = \Lambda$$

$$\Delta_g = \Delta$$

$$\Delta_g = I$$

$$\omega_g = \omega$$

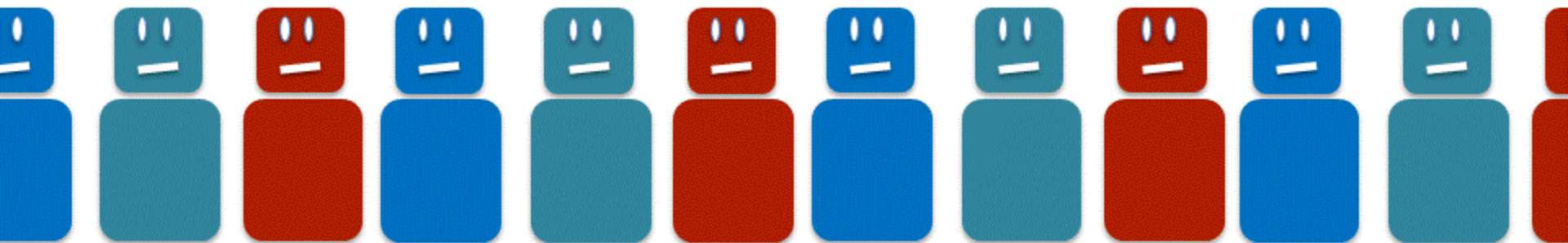
Model selection via the  
**Bayesian Information Criterion (BIC)**,  
which imposes a penalty for each  
additional parameter.



Clusters: heterogeneous

Products: variables

Order: nil



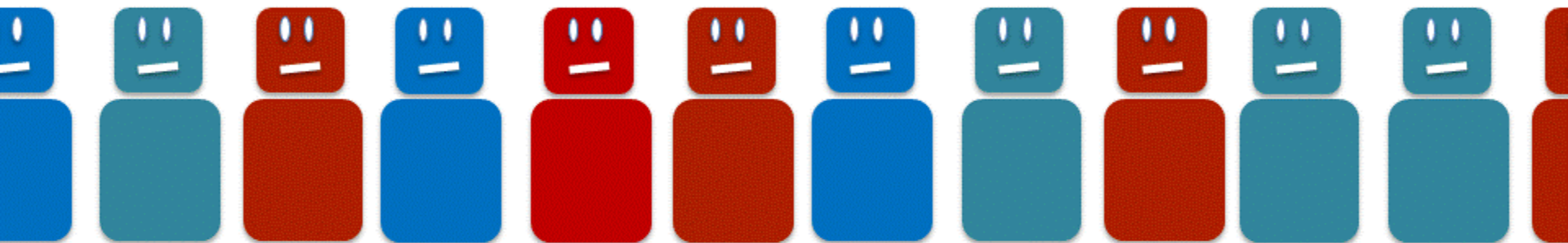
Obtain the best model\*

\* e.g. Franczak et al. (2015) used a mixture of factor analyzers with data imputation that was updated iteratively based on predicted cluster memberships

Clusters: heterogeneous

Products: variables

Order: homogeneous



Estimate and remove *common* order effects

Penalize BIC accordingly.

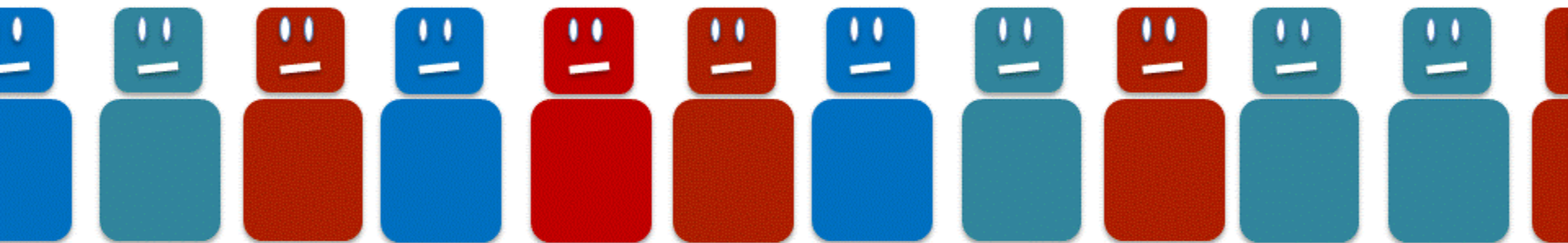
Obtain best mixture model.

Research in Progress...

Clusters: heterogeneous

Products: variables

Order: heterogeneous



Estimate & remove order effects *per group*.

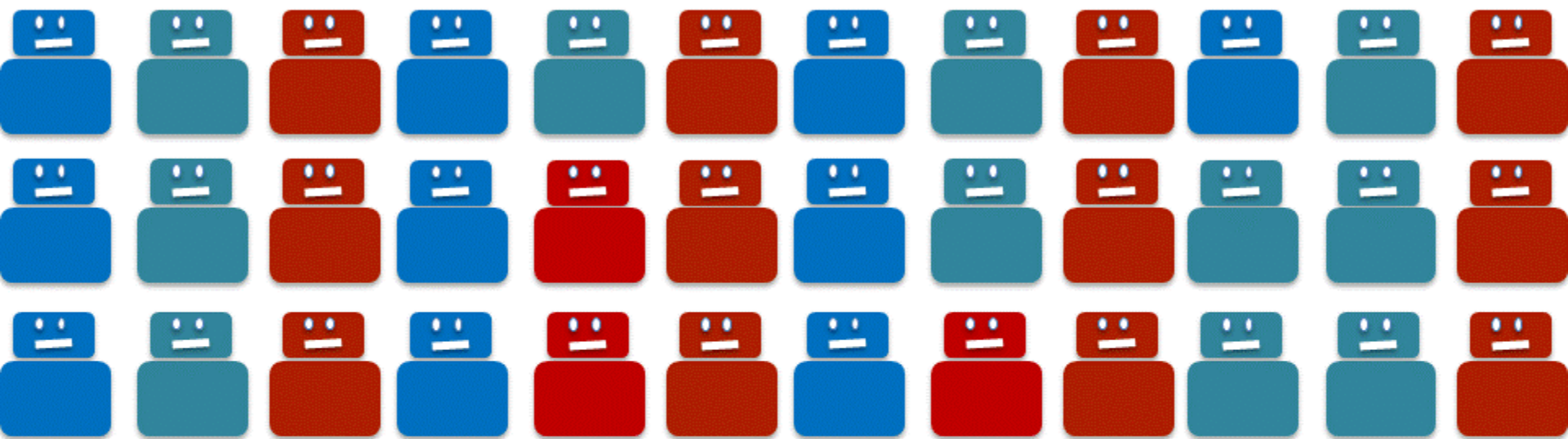
Penalize BIC accordingly.

Obtain best mixture model.

Research in Progress...



# Comparison



Model selection via BIC.

Review plots (original scale and latent space).

Research in Progress...





**Consumer acceptance  
(repeated measures)**



# Some Potential Strategies

## Conventional clustering

use summary data (e.g. sample means)  
unfold data

## Clustering matrices

E.g., cluster consumers assuming a  
mixture of matrix normal distributions

		1 min	2 min	5 min	10 min
A	4 <sup>th</sup>	7	6	6	5
B	3 <sup>rd</sup>	5	7	6	4
C	1 <sup>st</sup>	8	7	6	5
D	2 <sup>nd</sup>	7	8	6	5



		2 min	5 min	10 min
A	1 <sup>st</sup>	8	8	8
B	4 <sup>th</sup>	5	5	5
C	3 <sup>rd</sup>	8	7	6
D	2 <sup>nd</sup>	8	6	6



		1 min	2 min	5 min	10 min
A	2 <sup>nd</sup>	7	7	8	7
B	1 <sup>st</sup>	9	8	8	6
C	4 <sup>th</sup>	5	6	7	6
D	3 <sup>rd</sup>	8	8	7	7





My responses are honest and momentary.



My liking responses are an honest integration of my response from my initial impressions until the current time.



My liking responses are an honest integration of my response since the last time I was asked.



To be consistent I give the sample the same liking response every time that I am asked.







I'm so happy to be here that I rate every sample as "Like Very Much"!







# Point in time data

e.g. at 10 min.





	A	B	C	D
	5	4	5	5
	8	5	6	6
	7	6	6	7
	5	5	8	5

# Summarize data

e.g. area under curve

	A	B	C	D
	24	22	26	26
	32	21	28	28
	29	31	24	30
	35	18	33	33

# Unfold data

	A1	A2	A5	A10	B1	B2	B5	B10	C1	C2	C5	C10	D1	D2	D5	D10
	7	6	6	5	5	7	6	4	8	7	6	5	7	8	6	5
	8	8	8	8	6	5	5	5	7	8	7	6	8	8	6	6
	7	7	8	7	9	8	8	6	5	6	7	6	8	8	7	7
	6	6	7	6	5	5	4	4	8	8	7	7	7	5	5	5



# Matrix clustering

This is a potential application for matrix normal mixture model-based clustering\*.

		1 min	2 min	5 min	10 min
A	4 <sup>th</sup>	7	6	6	5
B	3 <sup>rd</sup>	5	7	6	4
C	1 <sup>st</sup>	8	7	6	5
D	2 <sup>nd</sup>	7	8	6	5



		1 min	2 min	5 min	10 min
A	1 <sup>st</sup>	8	8	8	8
B	4 <sup>th</sup>	6	5	5	5
C	3 <sup>rd</sup>	7	8	7	6
D	2 <sup>nd</sup>	8	8	6	6



		1 min	2 min	5 min	10 min
A	2 <sup>nd</sup>	7	7	8	7
B	1 <sup>st</sup>	9	8	8	6
C	4 <sup>th</sup>	5	6	7	6
D	3 <sup>rd</sup>	8	8	7	7



\* See Li (2014, Ch. 3) for applications of matrix clustering to selected sensory evaluation data.



The image shows three loaves of bread against a white background. The top loaf is a standard white sandwich loaf, partially sliced. The middle loaf is a darker, seeded bread, also partially sliced. The bottom loaf is a very dark, heavily seeded bread, with several slices cut and arranged in front of it. Overlaid on the bread is the text 'Incomplete block designs' in a light blue, sans-serif font.

# Incomplete block designs

# Balanced Incomplete Block Design

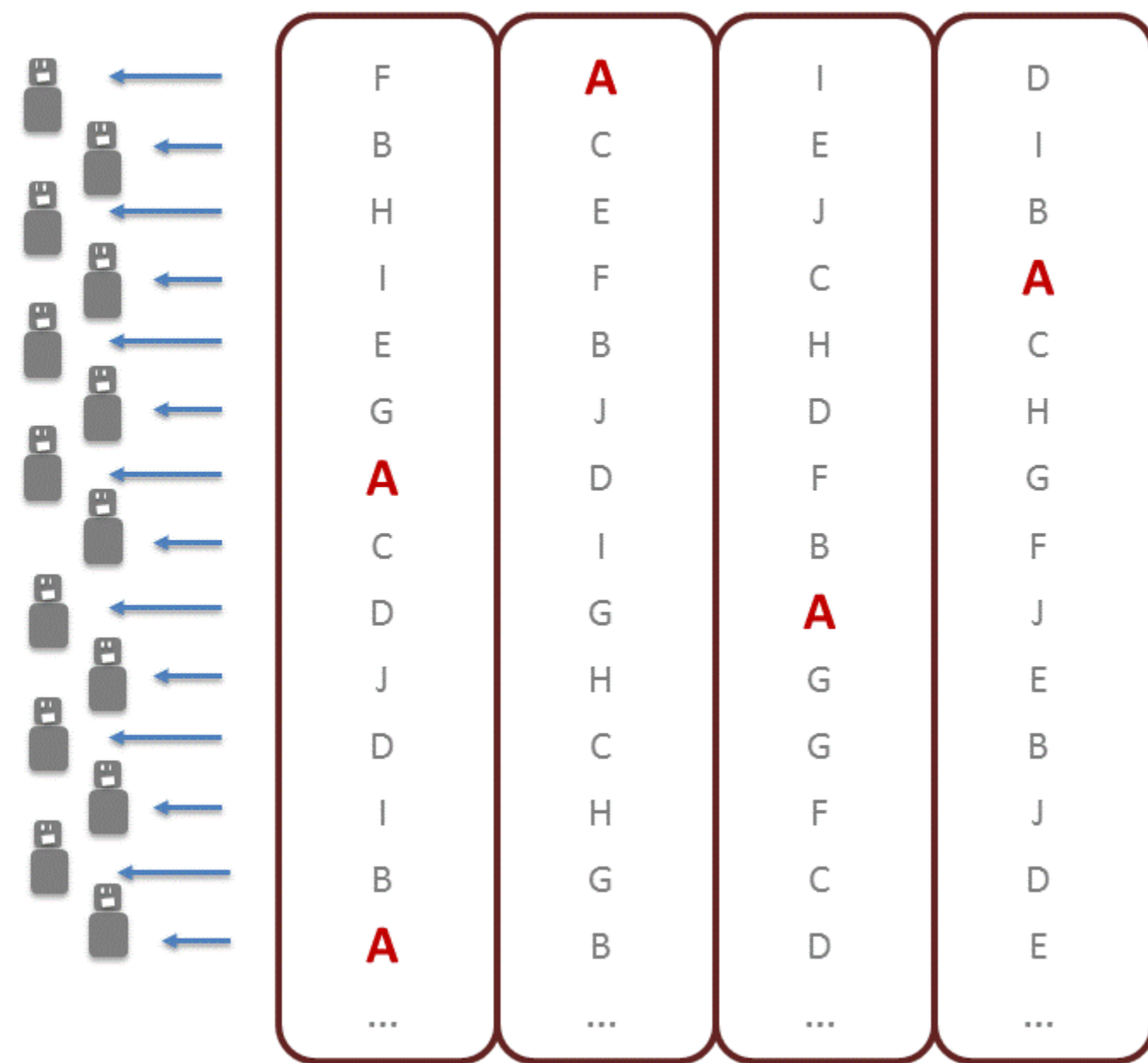
Each consumer evaluates  $k$  of  $t$  samples  
( $k < t$ )

$t$ -present- $k$  design

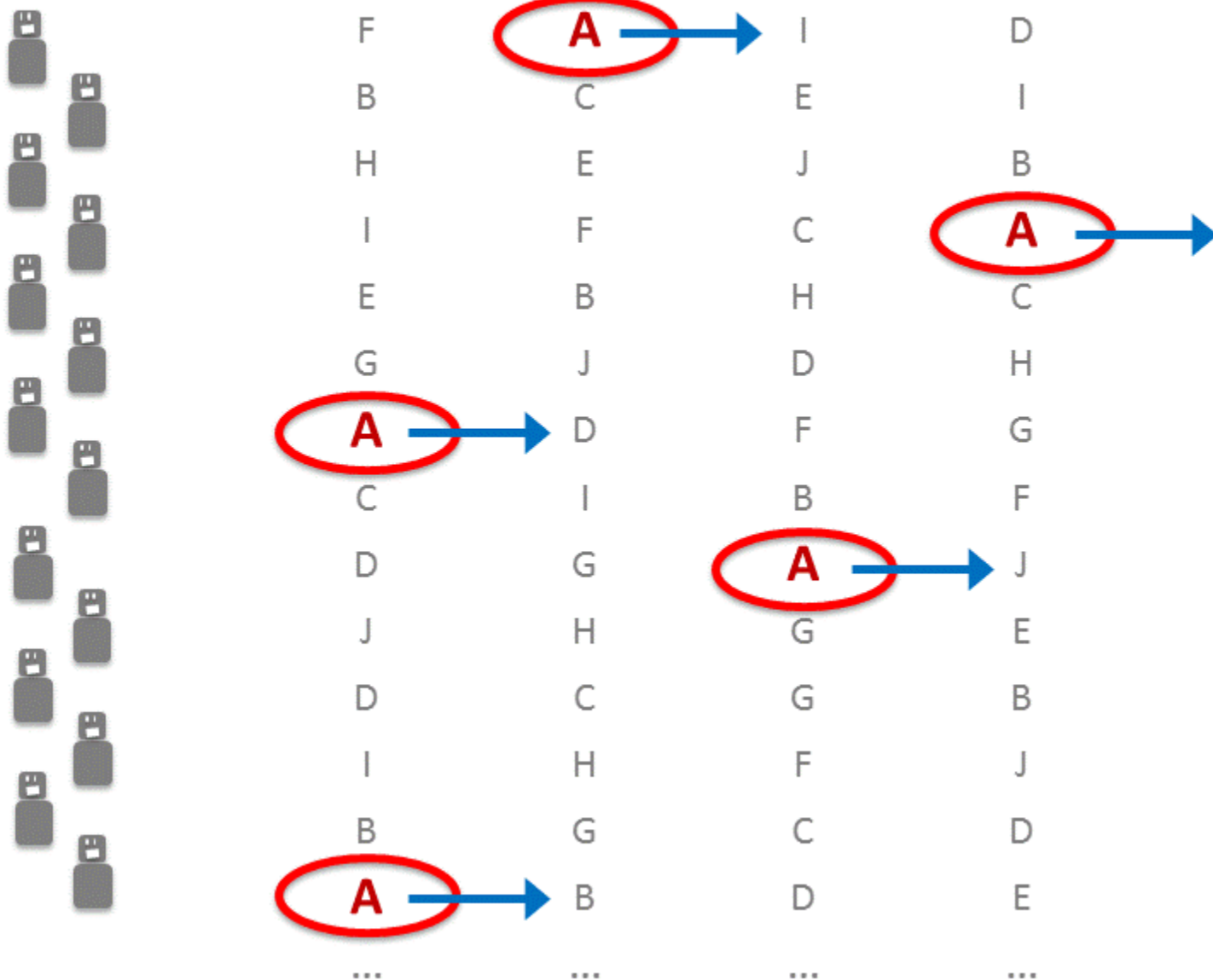
Goal:

Order and carryover balanced  
Unbiased product estimates

# Balanced Incomplete Block Design

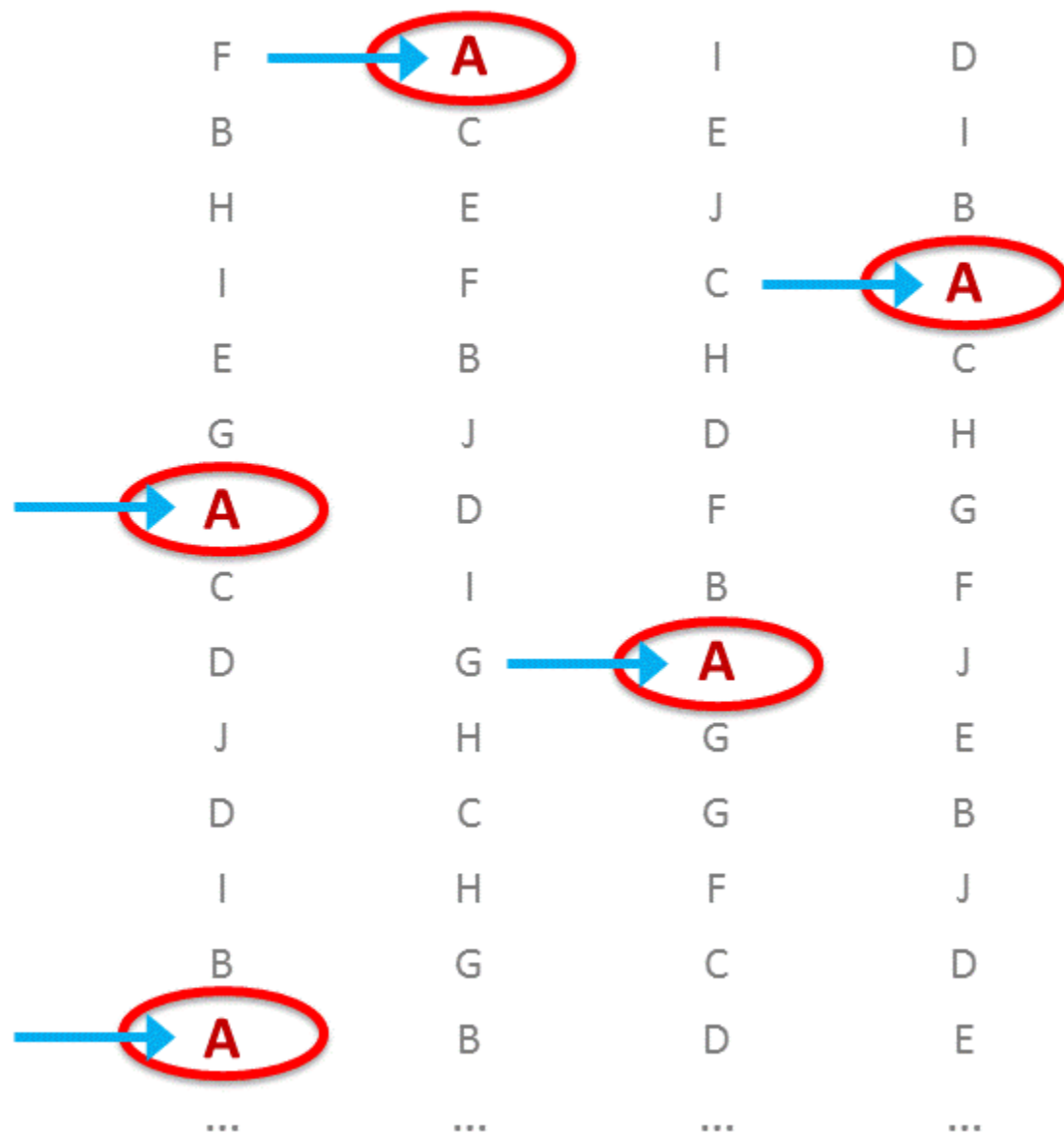


# Balanced Incomplete Block Design







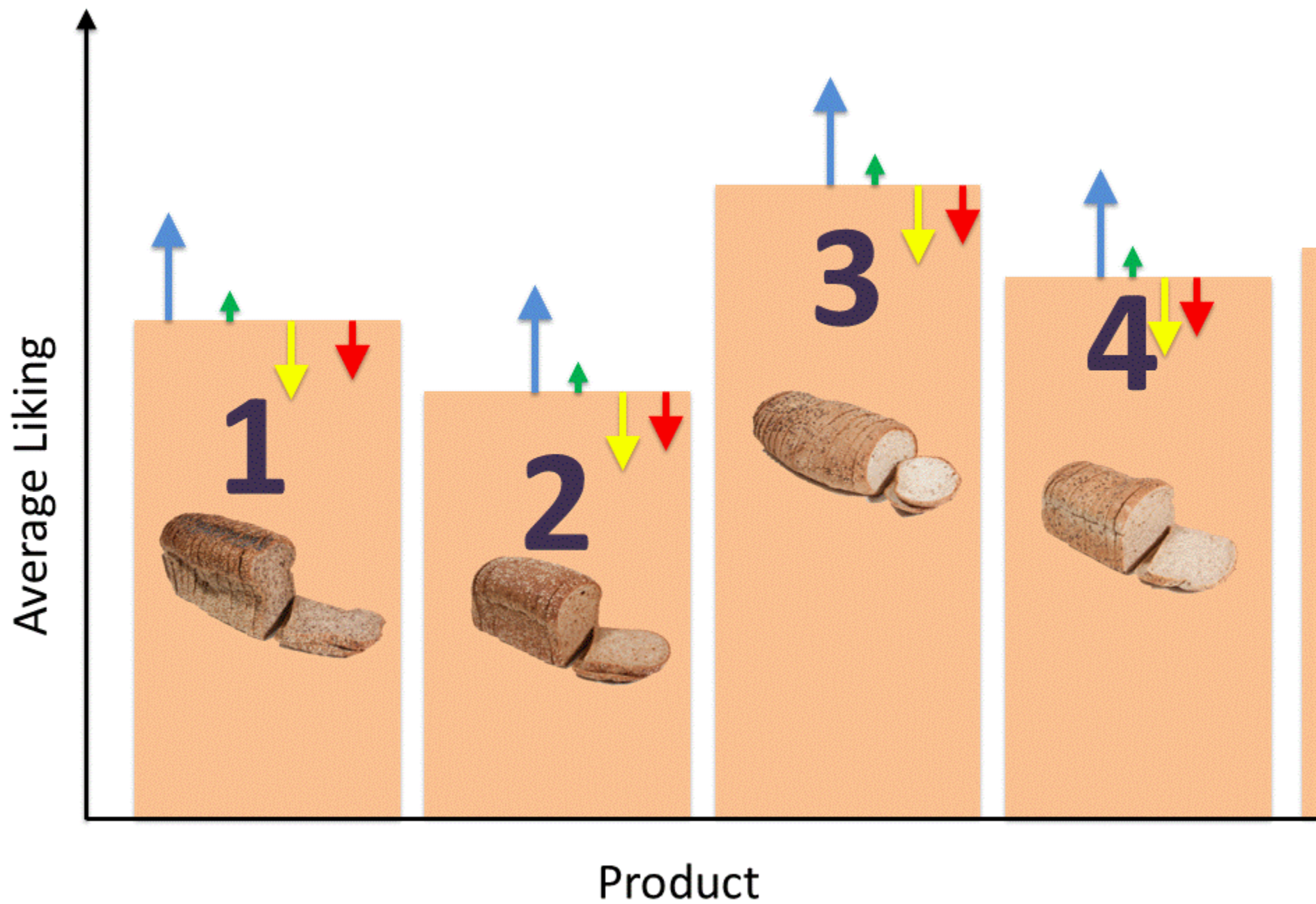


# Balanced Incomplete Block Design



## Consumer data

			6	5		7			5	
	8	8			4			5		
			8	7			8			6
		6				5	5			9

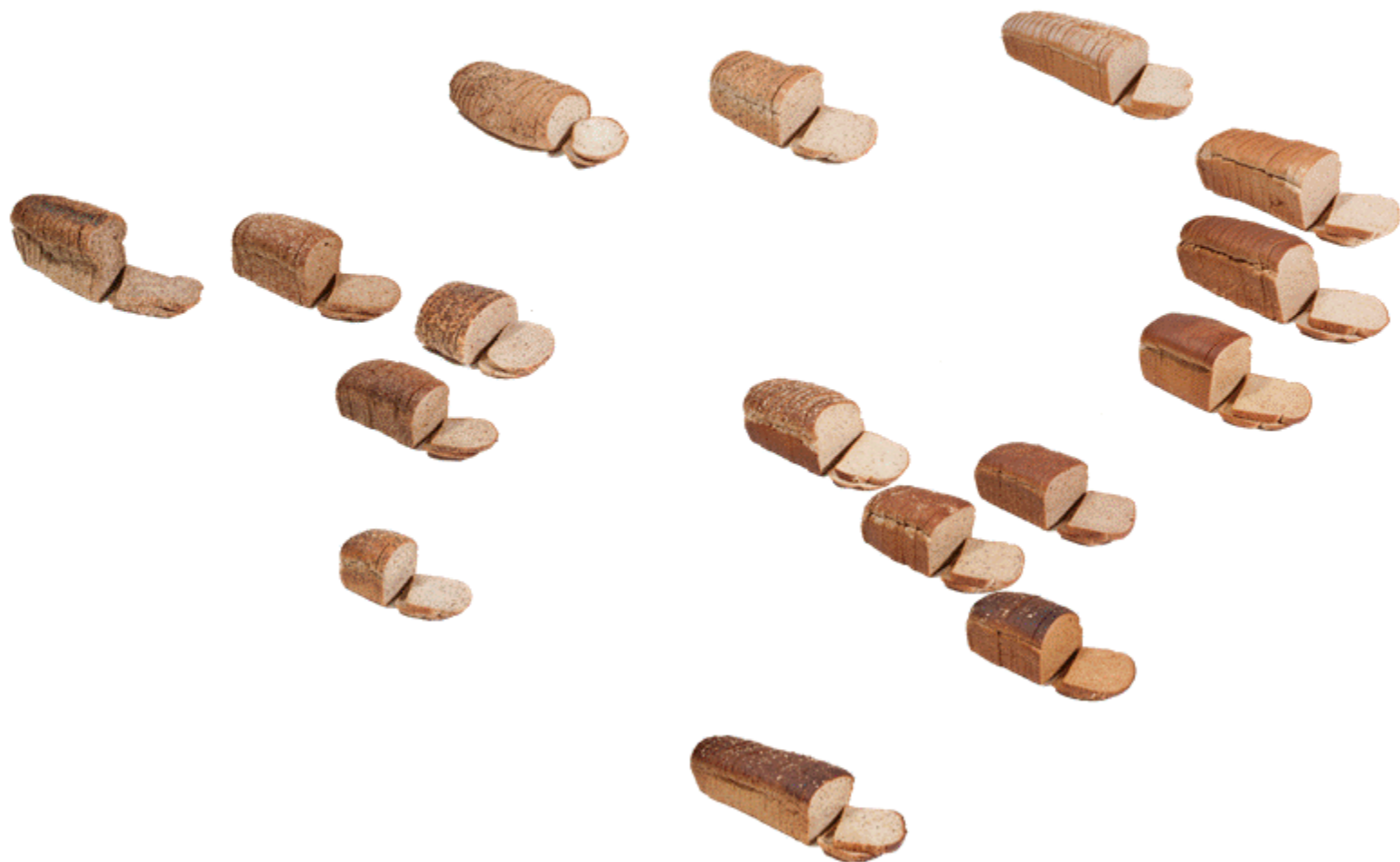


**There is additional information!**

**A trained sensory descriptive  
analysis panel evaluated 16  
whole grain breads...**

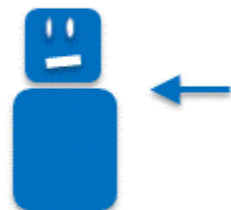


# Sensory space

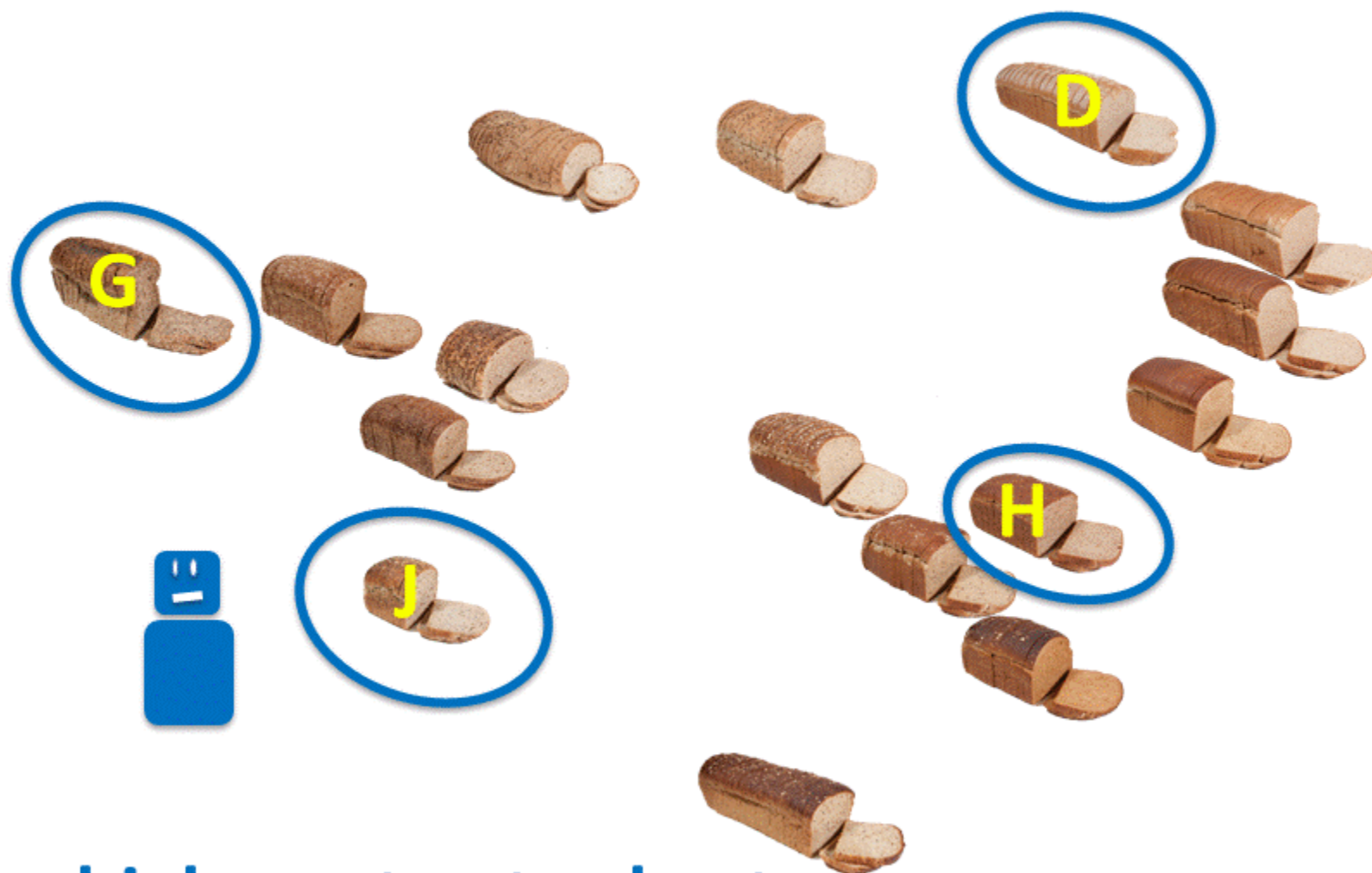


# Balanced Incomplete Block Design

F	A	I	D
B	C	E	I
H	E	J	B
I	F	C	A
E	B	H	C
<b>G</b>	<b>J</b>	<b>D</b>	<b>H</b>
A	D	F	G
C	I	B	F
D	G	A	J
J	H	G	E
D	C	G	B
I	H	F	J
B	G	C	D
A	B	D	E
...	...	...	...



## Sample set order #6



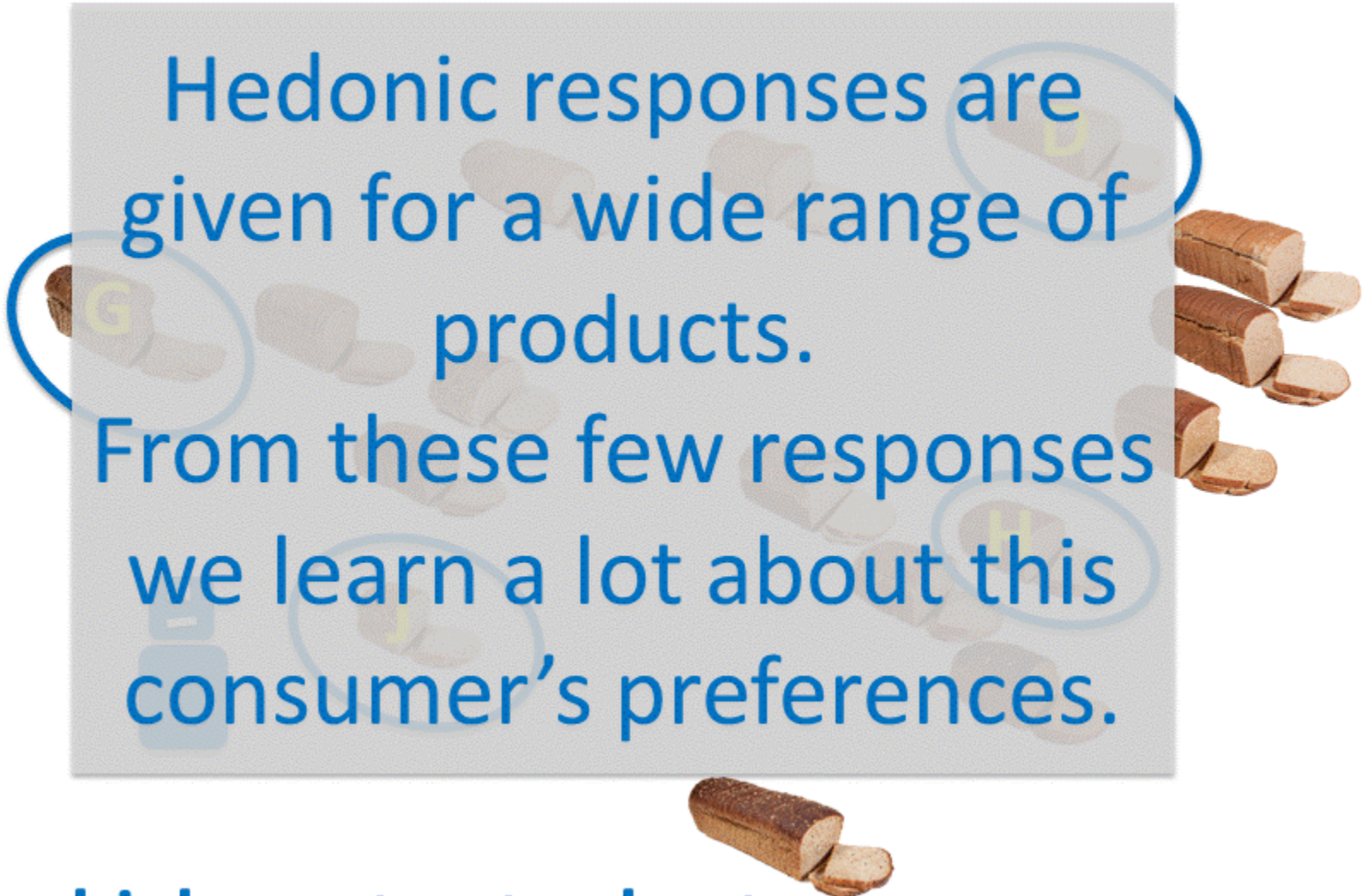
a high-contrast subset

## Sample set order #6

Hedonic responses are given for a wide range of products.

From these few responses we learn a lot about this consumer's preferences.

a high-contrast subset





# Balanced Incomplete Block Design

F	A	I	D
B	C	E	I
H	E	J	B
I	F	C	A
<b>E</b>	<b>B</b>	<b>H</b>	<b>C</b>
G	J	D	H
A	D	F	G
C	I	B	F
D	G	A	J
J	H	G	E
D	C	G	B
I	H	F	J
B	G	C	D
A	B	D	E
...	...	...	...



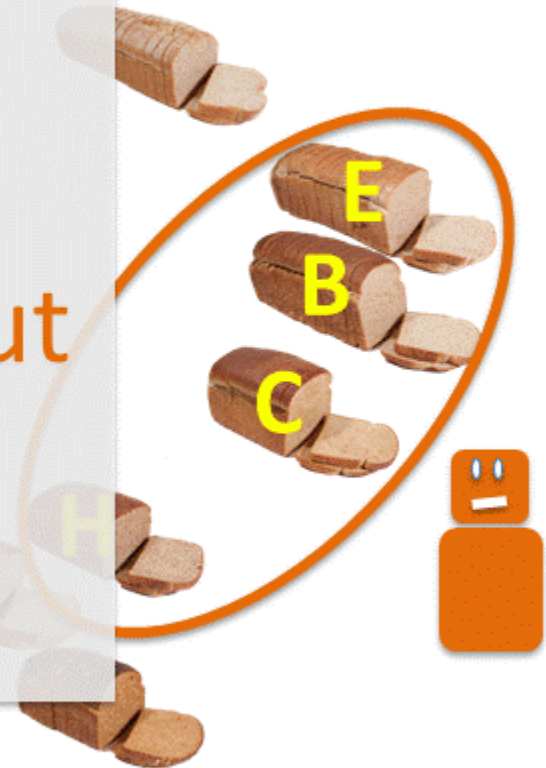
## Sample set order #5



a low-contrast subset

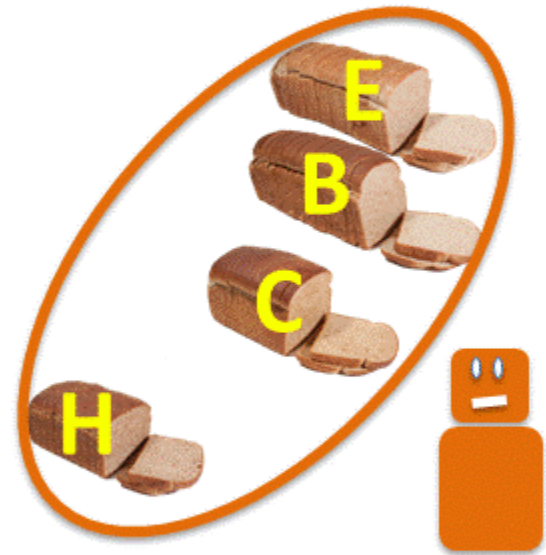
## Sample set order #5

Hedonic responses are given for a narrow range of products. So we learn little about this consumer's preferences.



a low-contrast subset

Would this  
product have  
been liked or  
disliked



a low-contrast subset



# Sensory Informed Design

$t$ -present- $k$  design

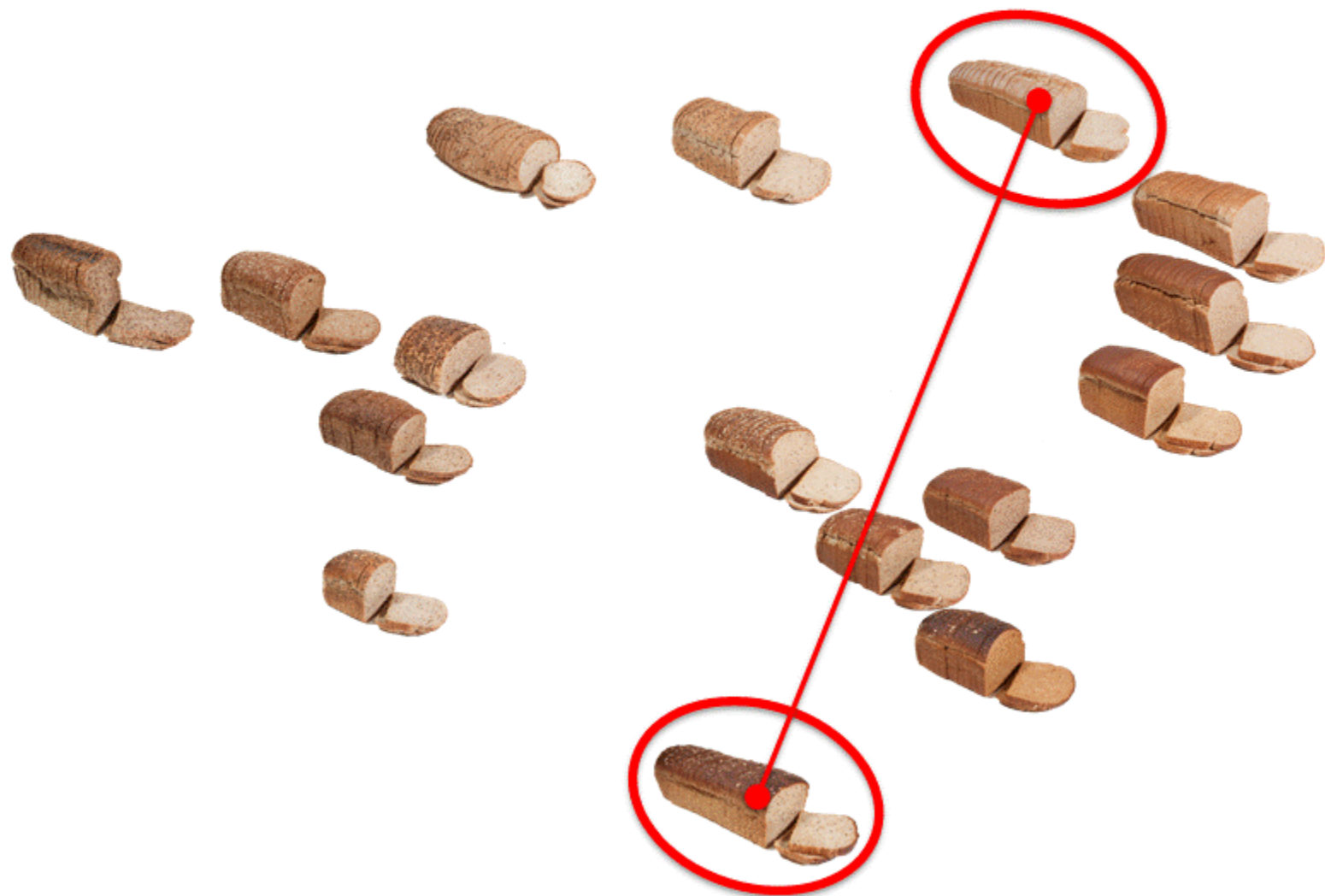
Goal:

Favour sample sets with **sensory contrast**

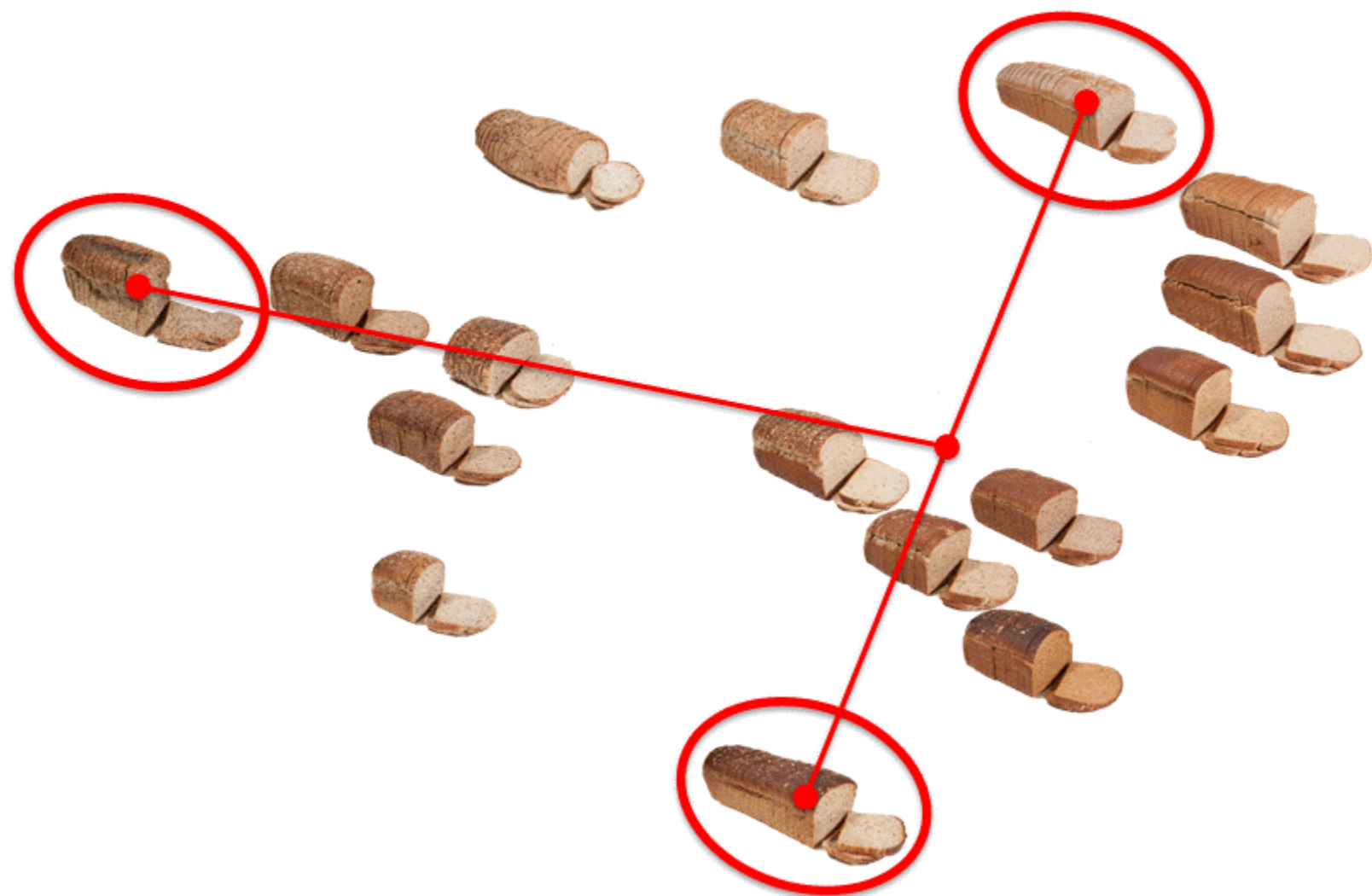
Order balanced

*Compromise:* carryover unbalanced

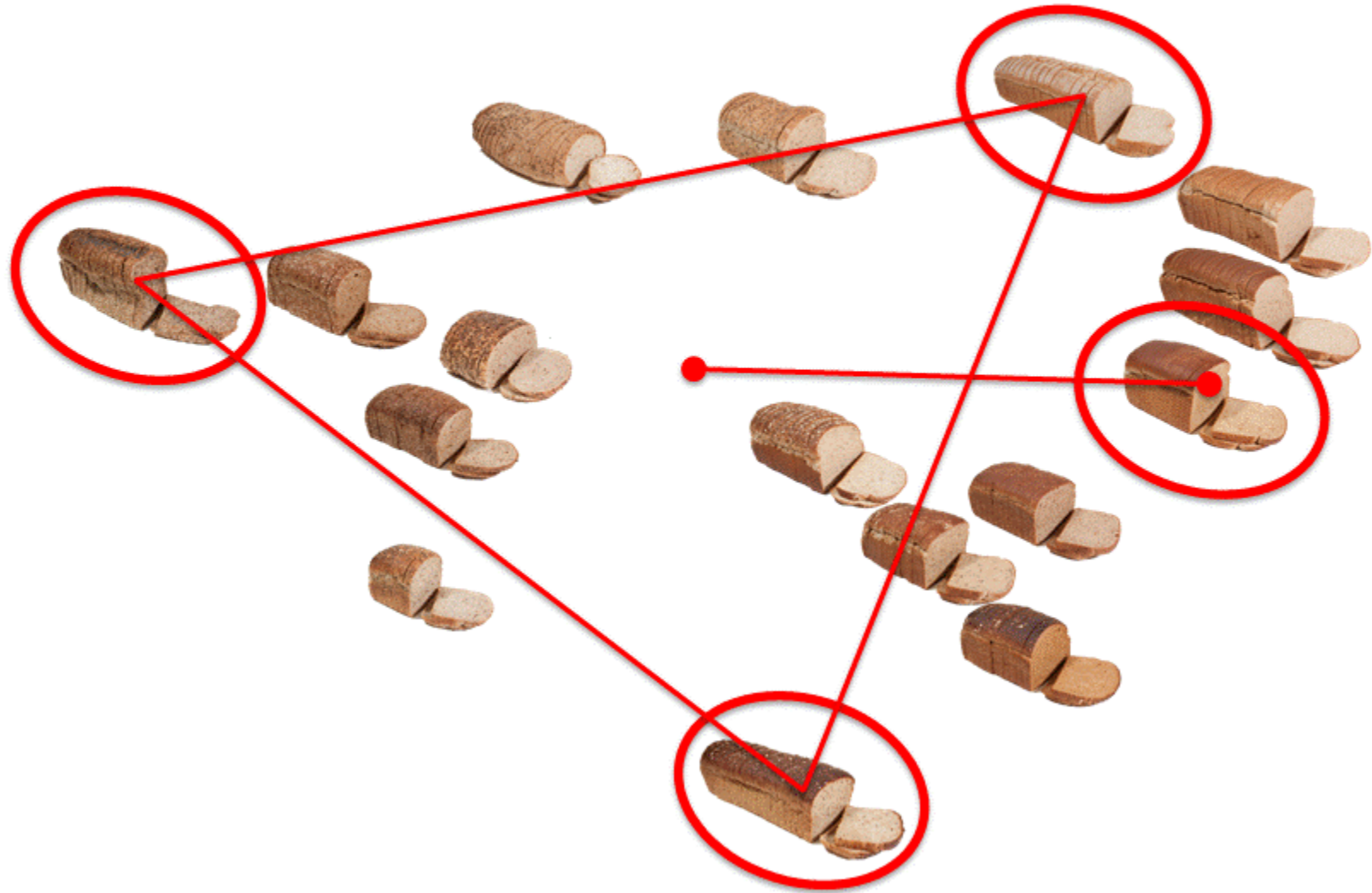
# Sensory Informed Design



# Sensory Informed Design

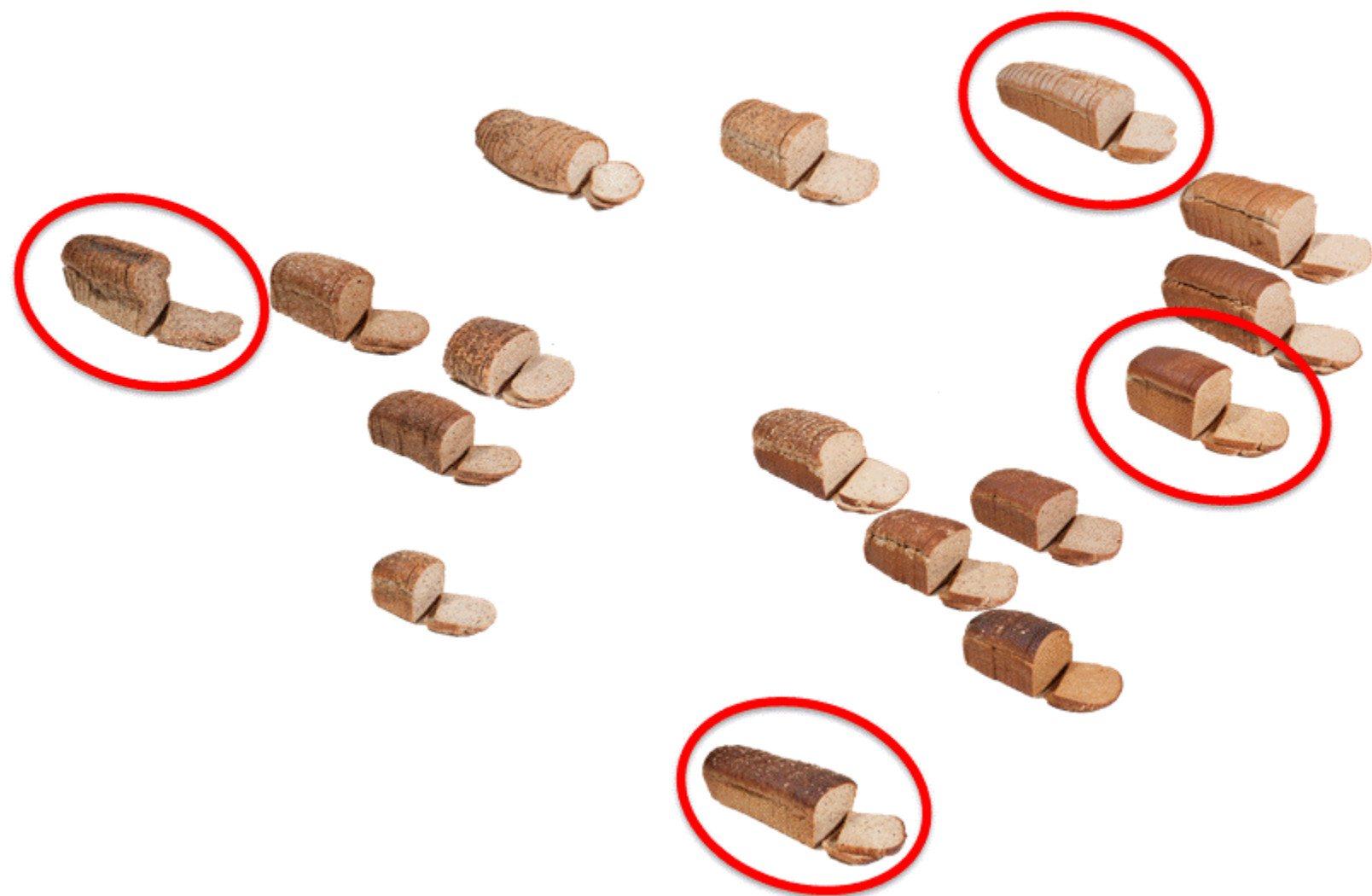


# Sensory Informed Design



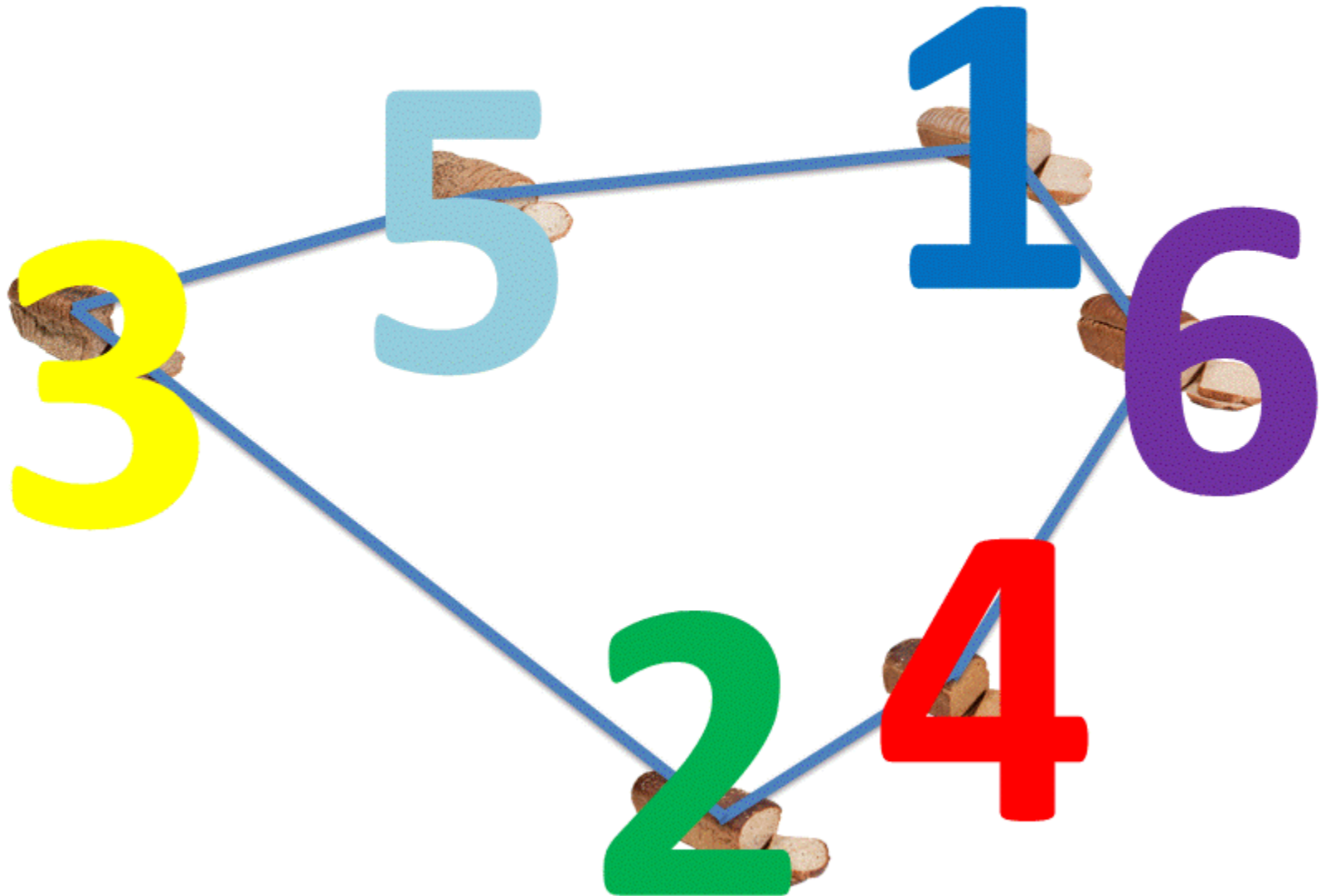


# Sensory Informed Design

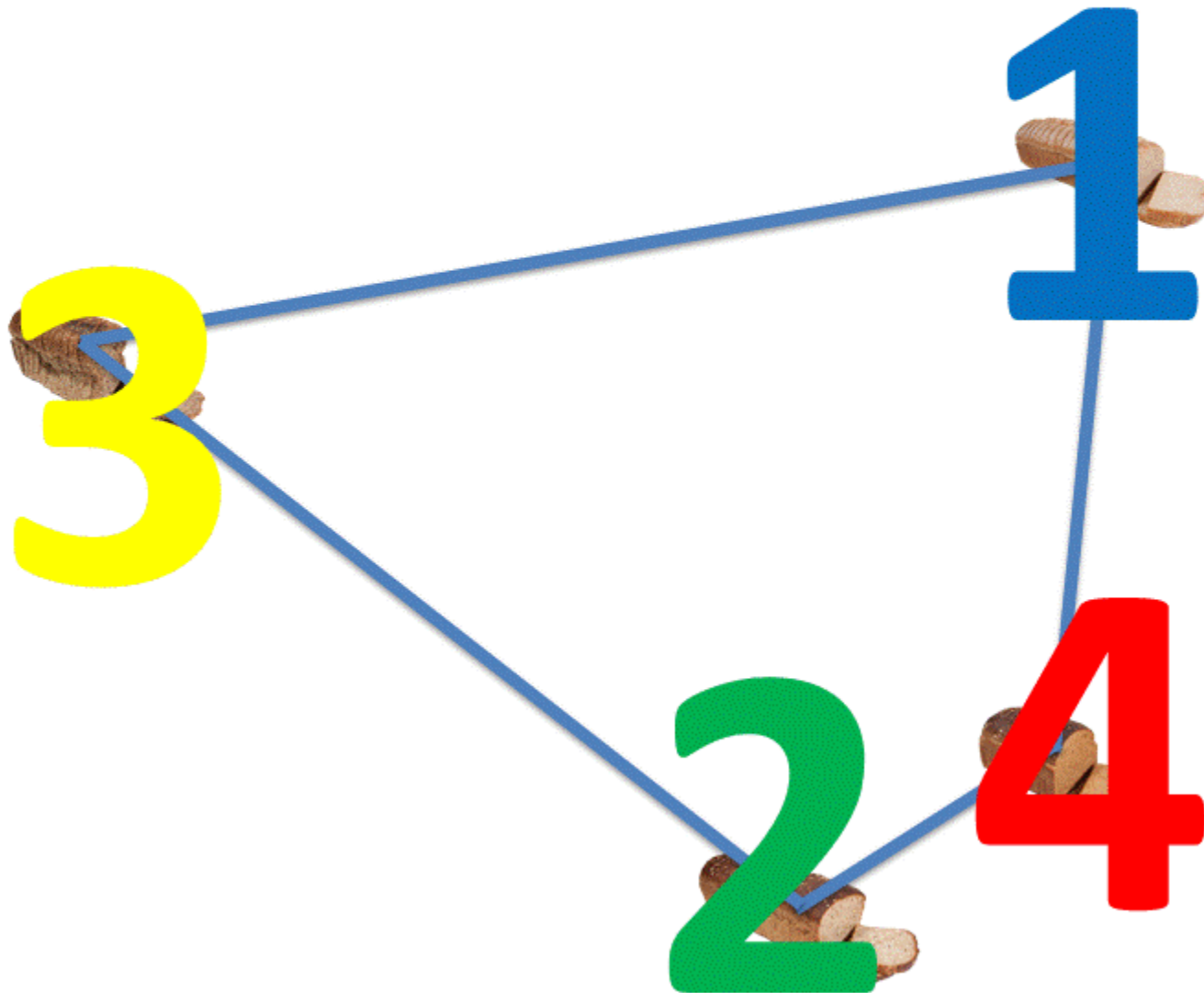


Franczak et al. (2015) describe a  
16-present-6 sensory informed design.

# Sensory Informed design (16-present-6 design)

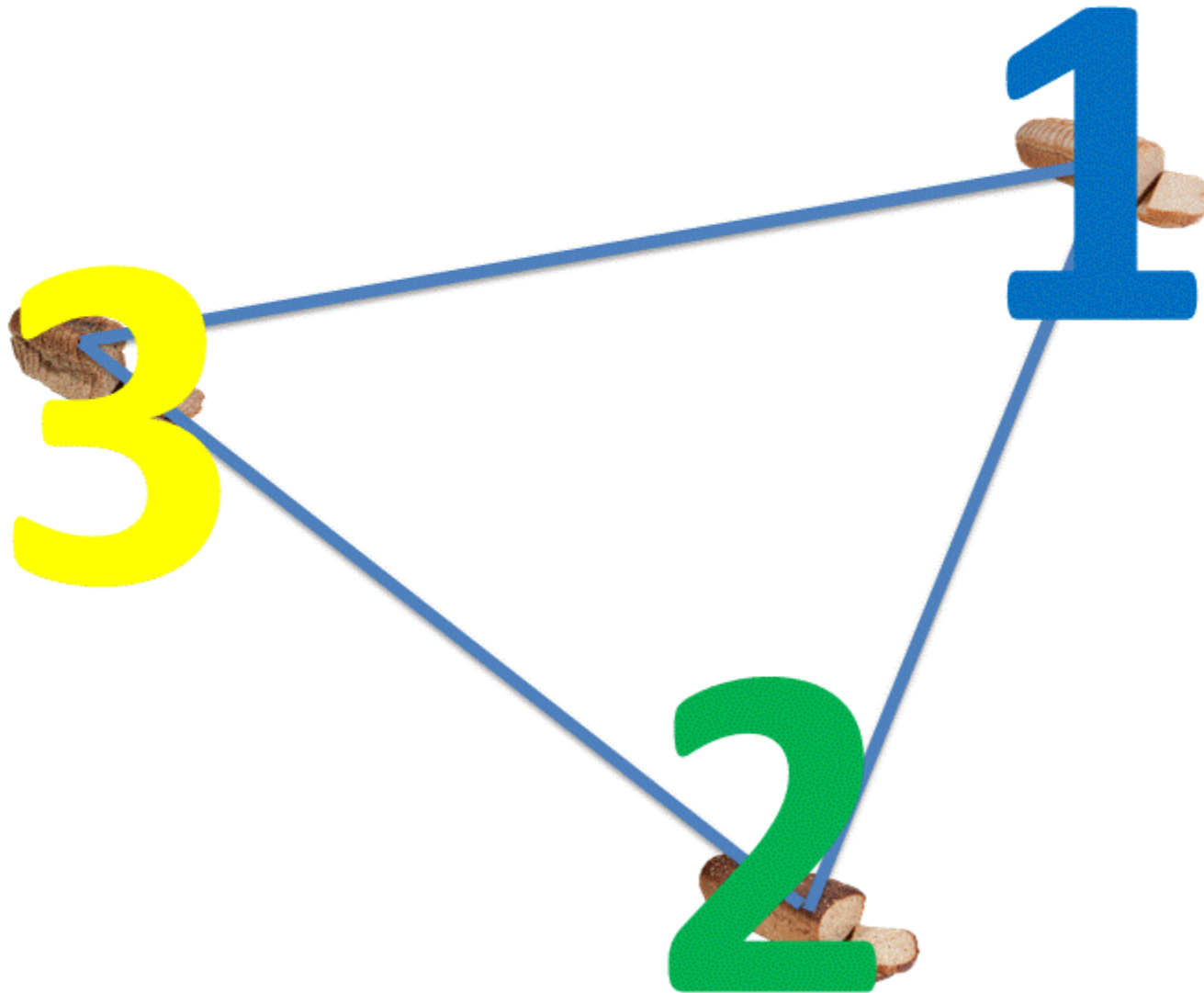


# Sensory Informed design (16-present-4 *nested* design)

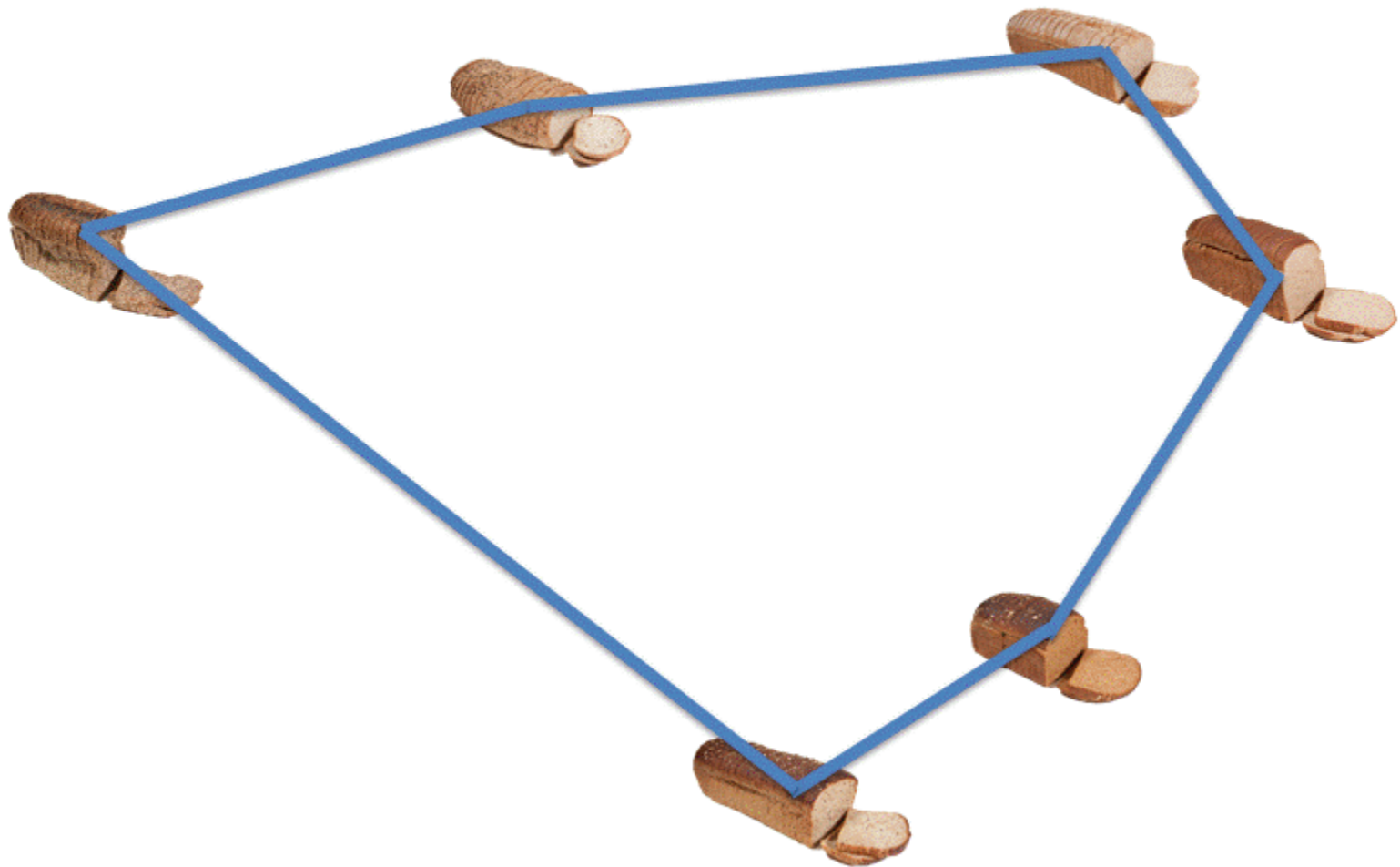




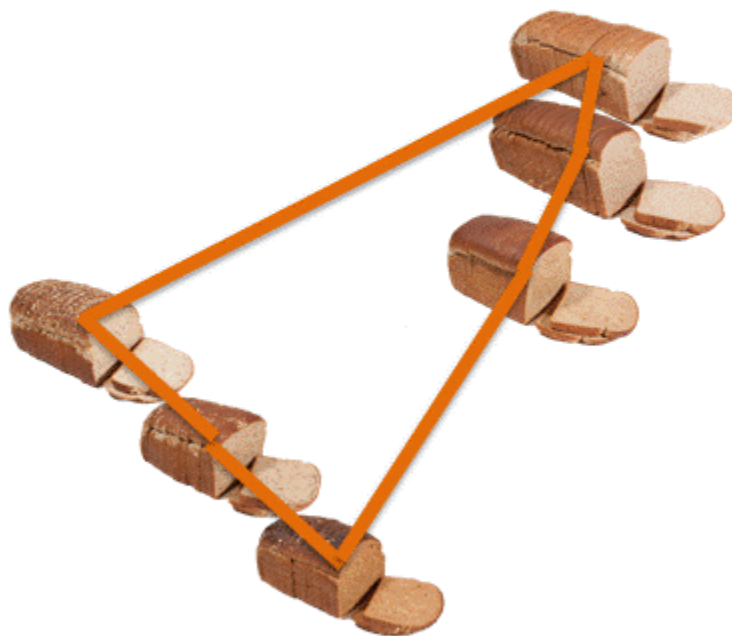
# Sensory Informed design (16-present-3 *nested* design)



# Scaling data?



# Scaling data?



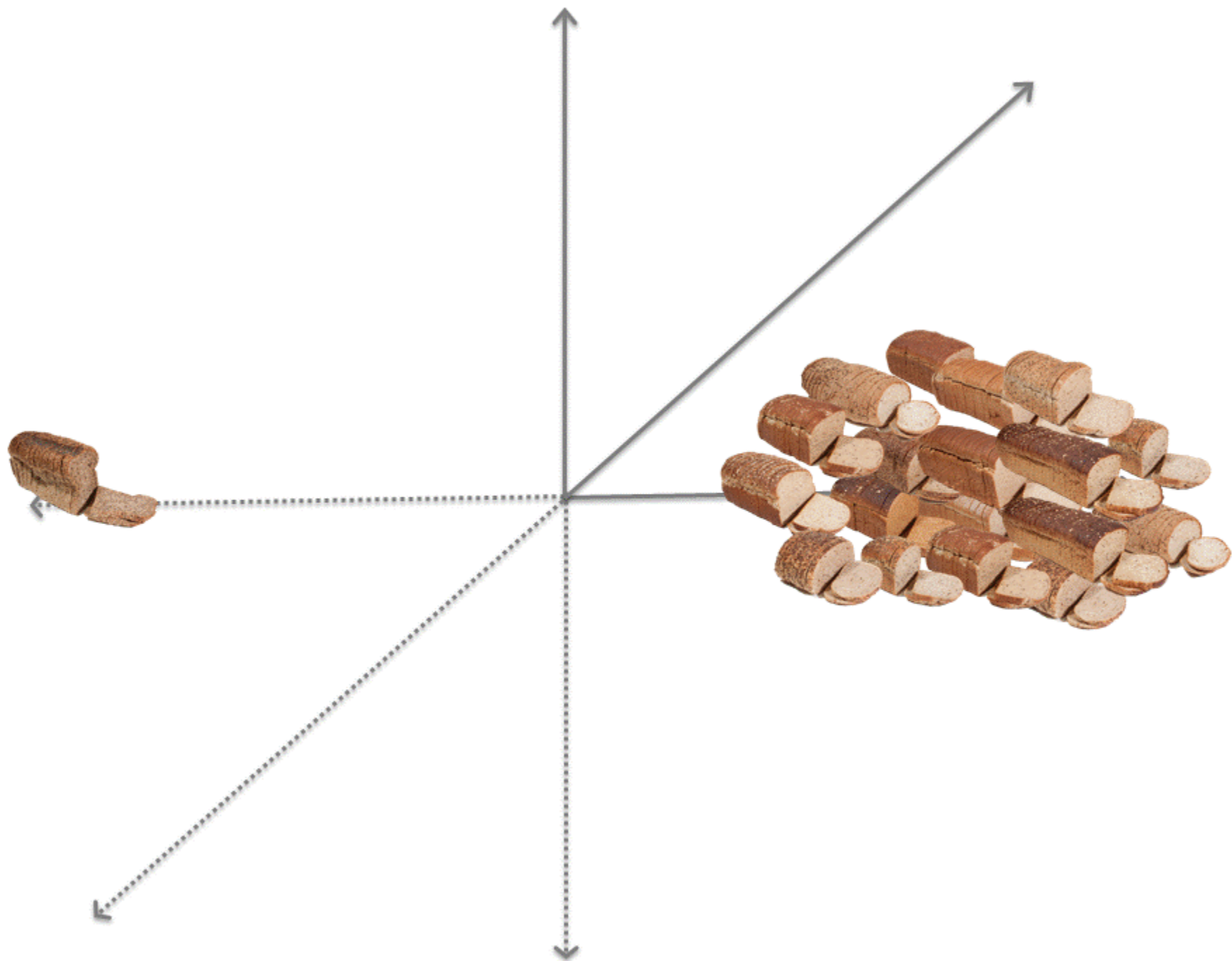
Is it *really* wise to  
center or rescale this  
consumer's liking data?

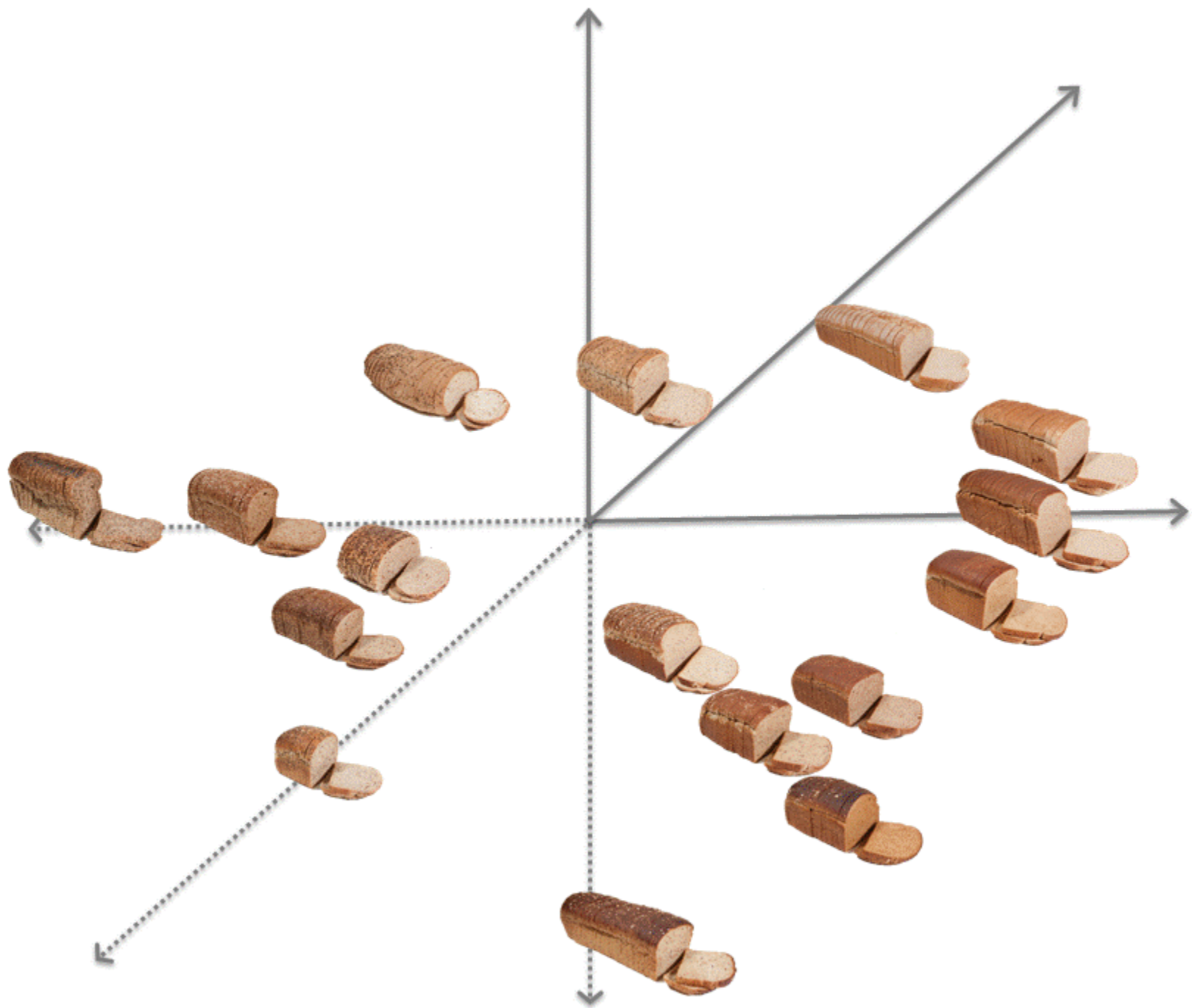
## Sensory Informed Designs

- t*** number of products  
product variability  
sensory space
- k*** number of samples presented
- N*** number of consumers  
consumer diversity
- & c.** context effects & biases (e.g. order)  
scale used for data collection  
sensory-liking relationship

***Further research required!***



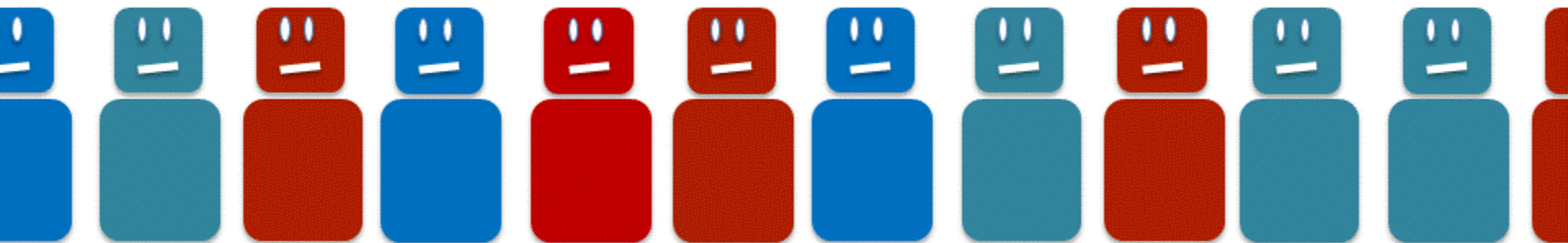




Clusters: heterogeneous

Products: variables

Order: heterogeneous



Estimate & remove order effects *per group*.

Estimate and impute missing data *per group* via conditional E-steps.

Obtain best mixture model

Penalize BIC according

Research in Progress...



# Part II: Consumer perception





**Please check all that apply.**

- |  |  |  |
|--|--|--|
| <input type="checkbox"/> Artificial Flavor           | <input type="checkbox"/> High Sweet Taste            | <input type="checkbox"/> Plastic Flavor                |
| <input type="checkbox"/> Bitter Taste                | <input type="checkbox"/> Lemon Flavor                | <input type="checkbox"/> Processed Flavor              |
| <input type="checkbox"/> Cheap Taste                 | <input type="checkbox"/> Low Sweet Taste             | <input type="checkbox"/> Refreshing Flavor             |
| <input type="checkbox"/> Earthy Flavor               | <input type="checkbox"/> Low Acidic/Sour/Tart Taste  | <input type="checkbox"/> Rotten/Overripe Orange Flavor |
| <input type="checkbox"/> Expensive Flavor            | <input type="checkbox"/> Natural Flavor              | <input type="checkbox"/> Shelf Stable Flavor           |
| <input type="checkbox"/> Fresh Orange Flavor         | <input type="checkbox"/> Not From Concentrate Flavor | <input type="checkbox"/> Strong Flavor                 |
| <input type="checkbox"/> Fresh Squeezed Flavor       | <input type="checkbox"/> Organic Flavor              | <input type="checkbox"/> Weak/Watery Flavor            |
| <input type="checkbox"/> From Concentrate Flavor     | <input type="checkbox"/> Other Citrus Flavor         | <input type="checkbox"/> None of these apply           |
| <input type="checkbox"/> Green/Unripe Orange Flavor  | <input type="checkbox"/> Oxidized Flavor             |  |
| <input type="checkbox"/> High Acidic/Sour/Tart Taste | <input type="checkbox"/> Papery/Cardboard Flavor     |  |

# Check-all-that-apply (CATA) questions

# Question order

**Liking → CATA**

Investigate

**perception responses** within  
**liking clusters**

and / or

**liking responses** within  
**perception clusters**

## Balance sample serving orders



## Balance attribute positions



A	B	D	C
B	C	A	D
C	D	B	A
D	A	C	B

B	A	C	D
A	D	B	C
D	C	A	B
C	B	D	A

C	B	A	D
B	D	C	A

a	b	h	c	g	d	f	e
b	c	a	d	h	e	g	f
c	d	b	e	a	f	h	g
d	e	c	f	b	e	a	h
e	f	d	g	c	f	b	a
f	g	e	h	d	g	c	b
g	h	f	a	e	a	d	c
h	a	g	b	f	b	e	d

g	e	c	d	a	b	h	d
e	d	a	b	c	f	a	h





**Rows: Consumers**

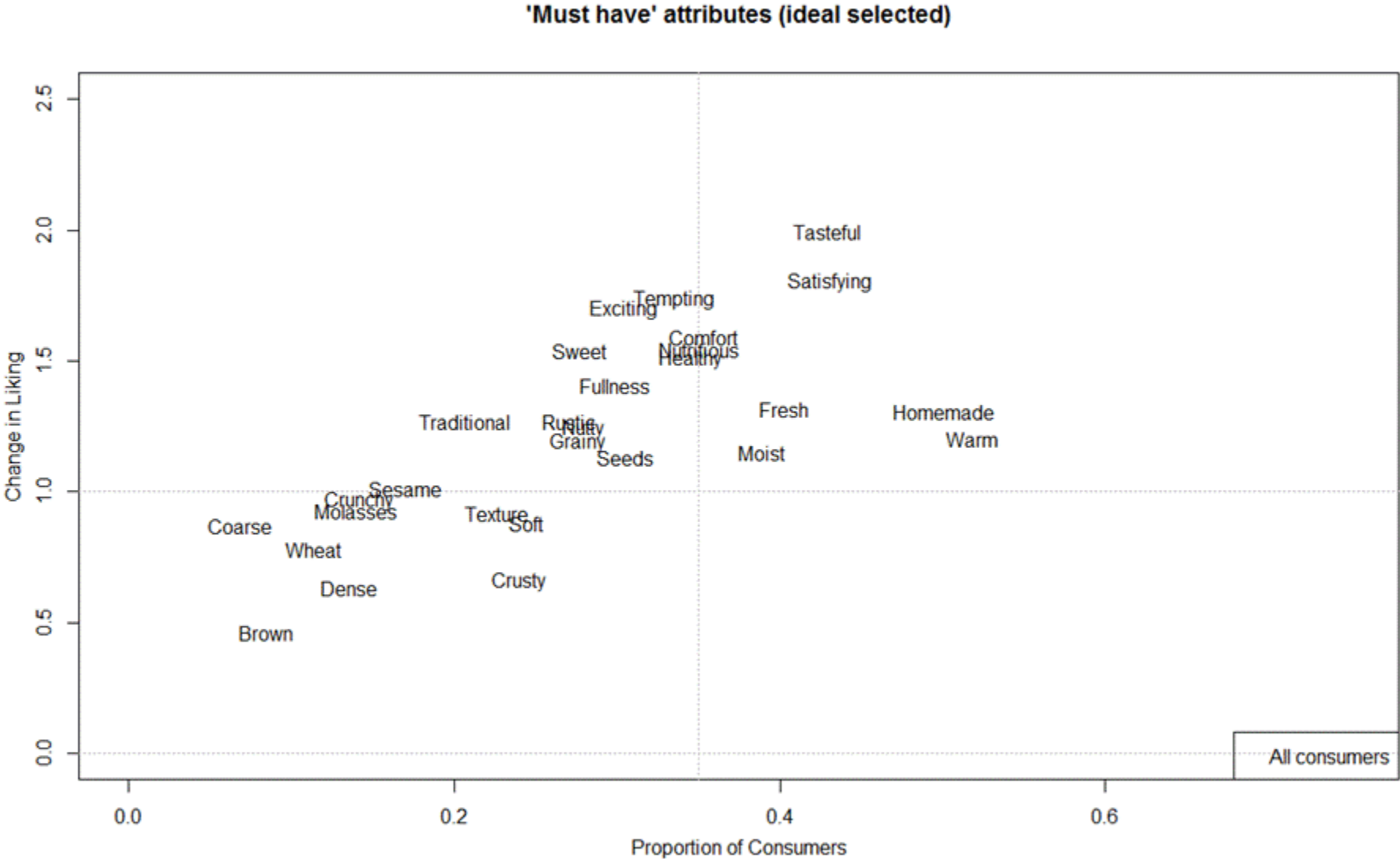
×

**Columns: Attributes**

×

**Slices: Products**

# Penalty analysis



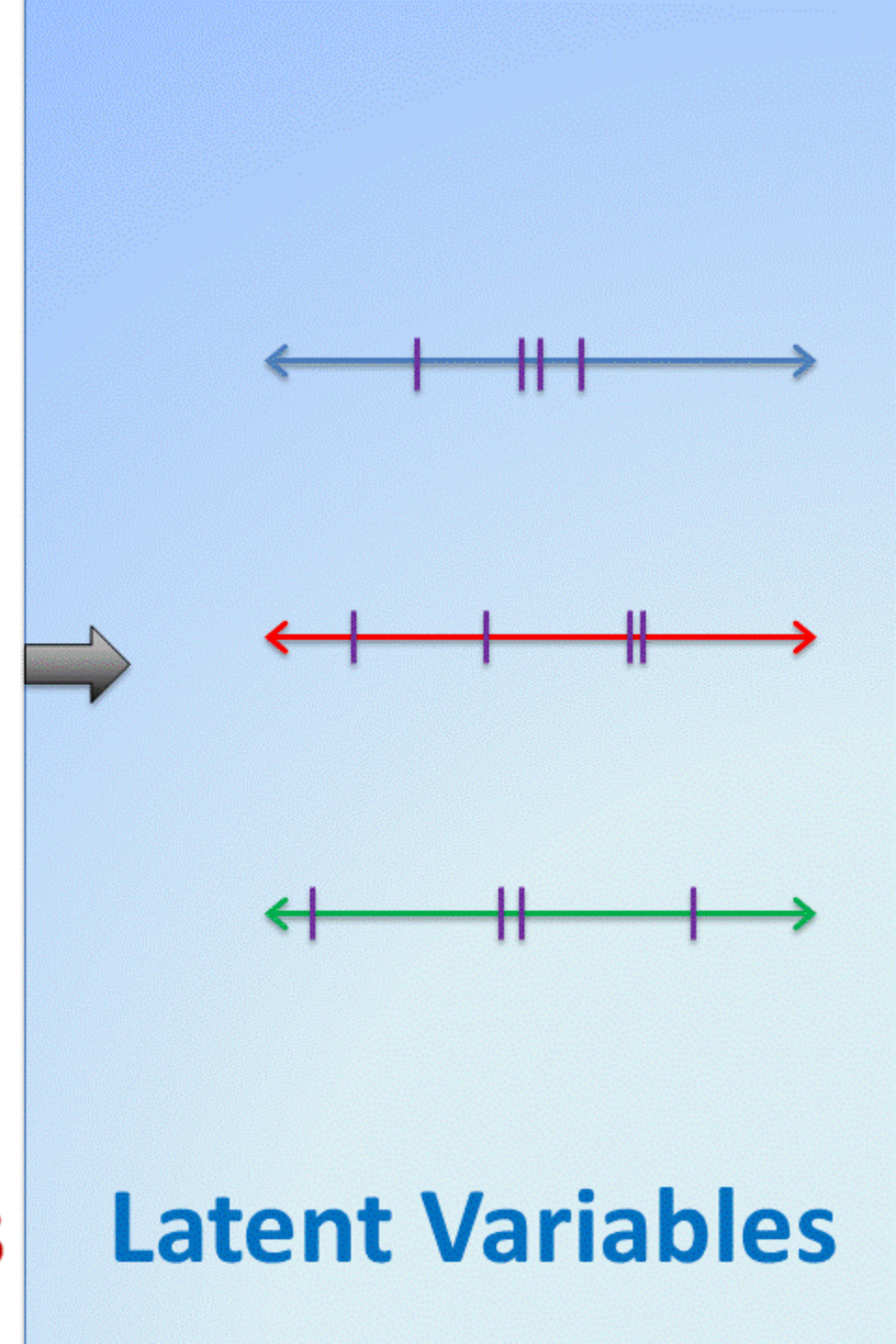
# “Ideal Product”

**Rows: Consumers**

×

**Columns: Attributes**







# Mixture of Latent Trait Models with Common Slope Parameters

Attribute  $k$ , Consumer  $i$ , Group  $g$

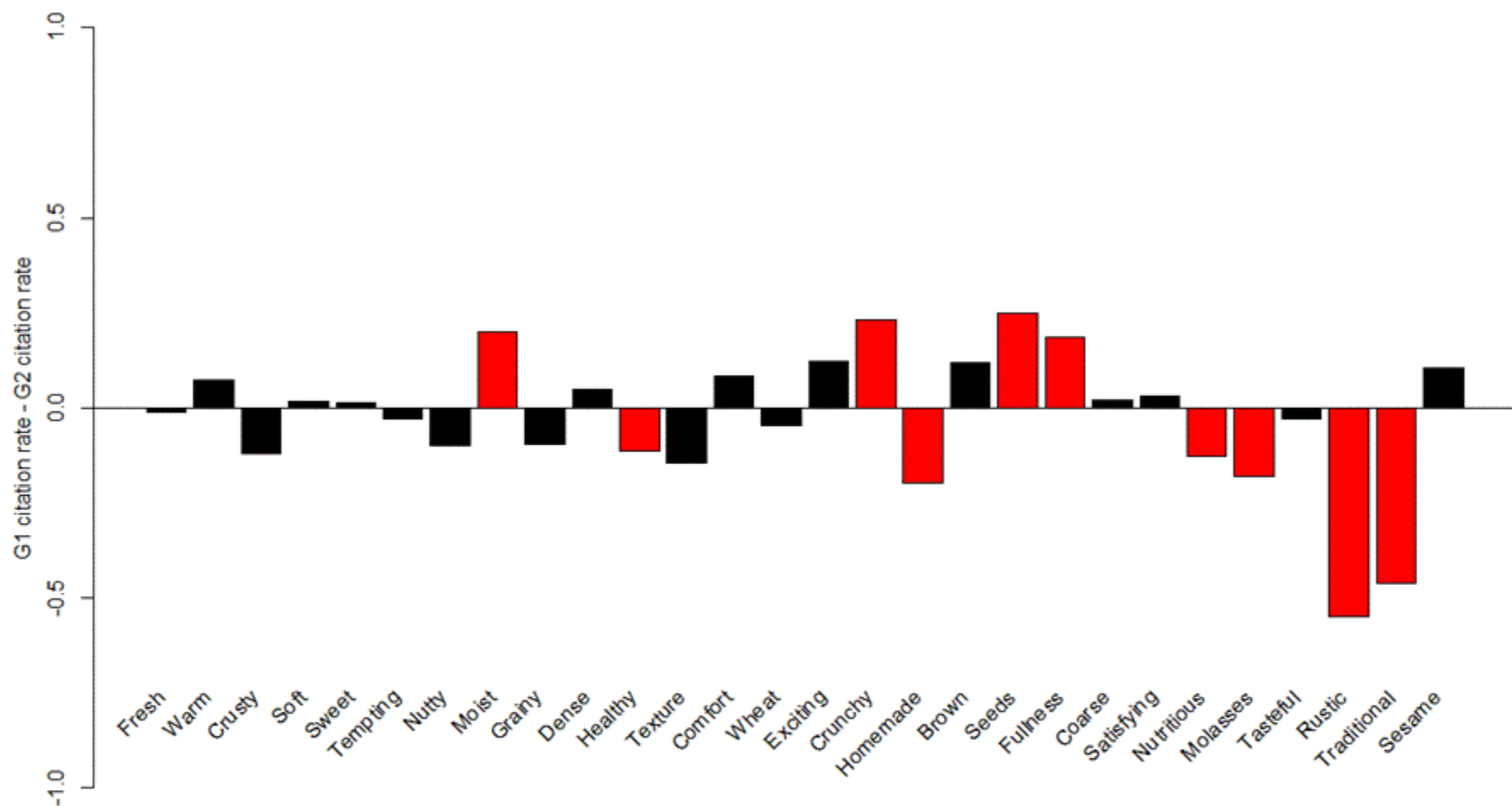
$$p(x_{ik} = 1 | y_{ig}, z_{ig} = 1) = \frac{1}{1 + \exp(-w'_k y_{ig})}$$

$$Y_{ig} \sim \text{MVN}(\mu_g, \Sigma_g)$$

$$\Sigma_g = \lambda_g Q_g A_g Q_g'$$

The diagram illustrates the decomposition of the covariance matrix  $\Sigma_g$  into three components: Volume, Orientation, and Shape. The equation  $\Sigma_g = \lambda_g Q_g A_g Q_g'$  is shown with colored terms:  $\lambda_g$  is blue,  $Q_g$  and  $Q_g'$  are orange, and  $A_g$  is purple. Below the equation, three labels are connected to the terms by lines: 'Volume' (blue) connects to  $\lambda_g$ , 'Orientation' (orange) connects to  $Q_g$ , and 'Shape' (purple) connects to  $A_g$ . There is also an orange line connecting  $Q_g'$  to the 'Orientation' label.

### Difference in Attribute Selection Rates for the Ideal Product



Consumers clustered by CATA profiles of an ideal bread into G1 (n=56) and G2 (n=105) via mixture of latent trait models with common slope parameters (MCLT)

BIC selects the following solution:

**2 groups, 2 latent variables**

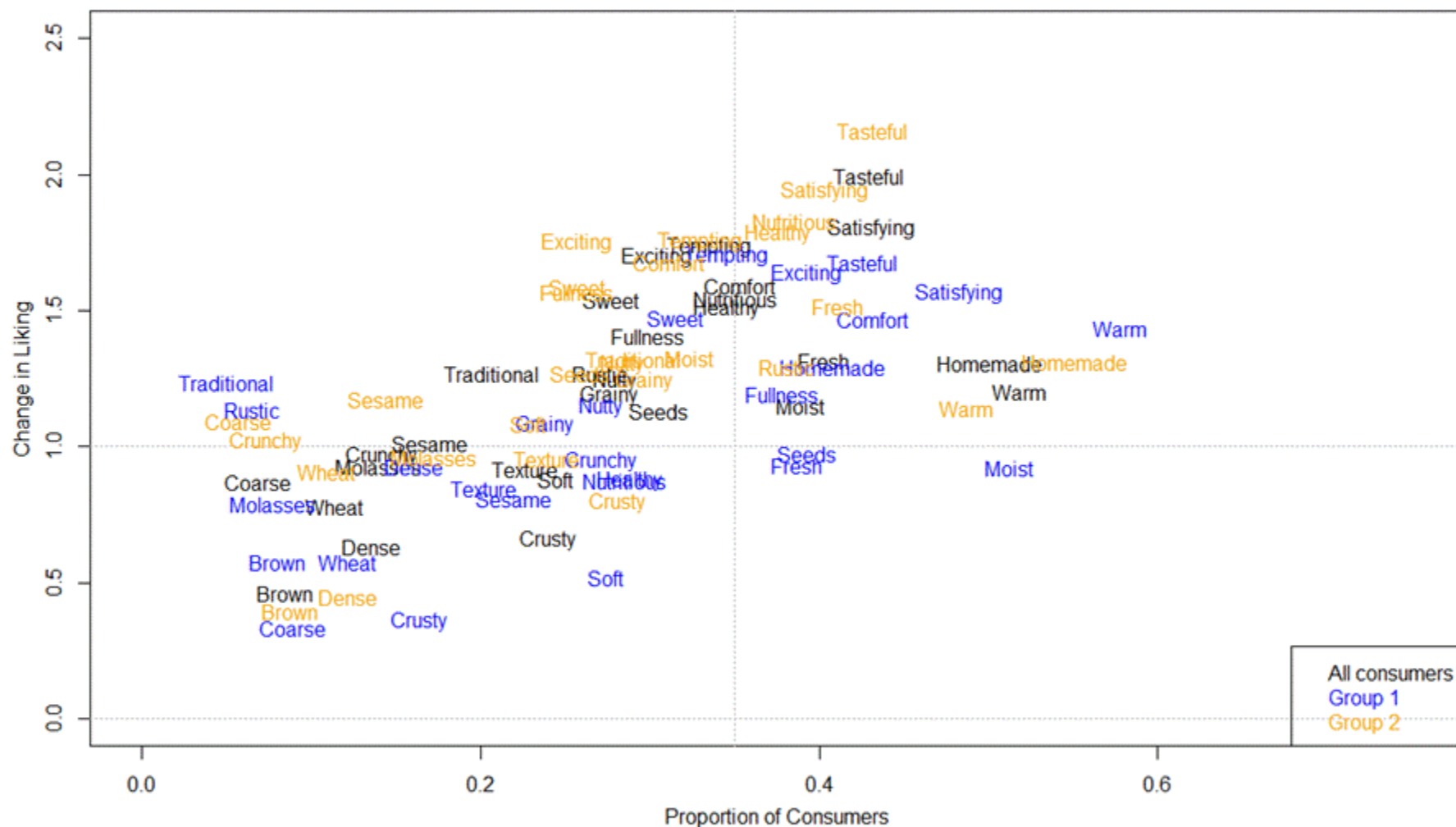


Citation proportions differ significantly

Citation proportions do not differ significantly

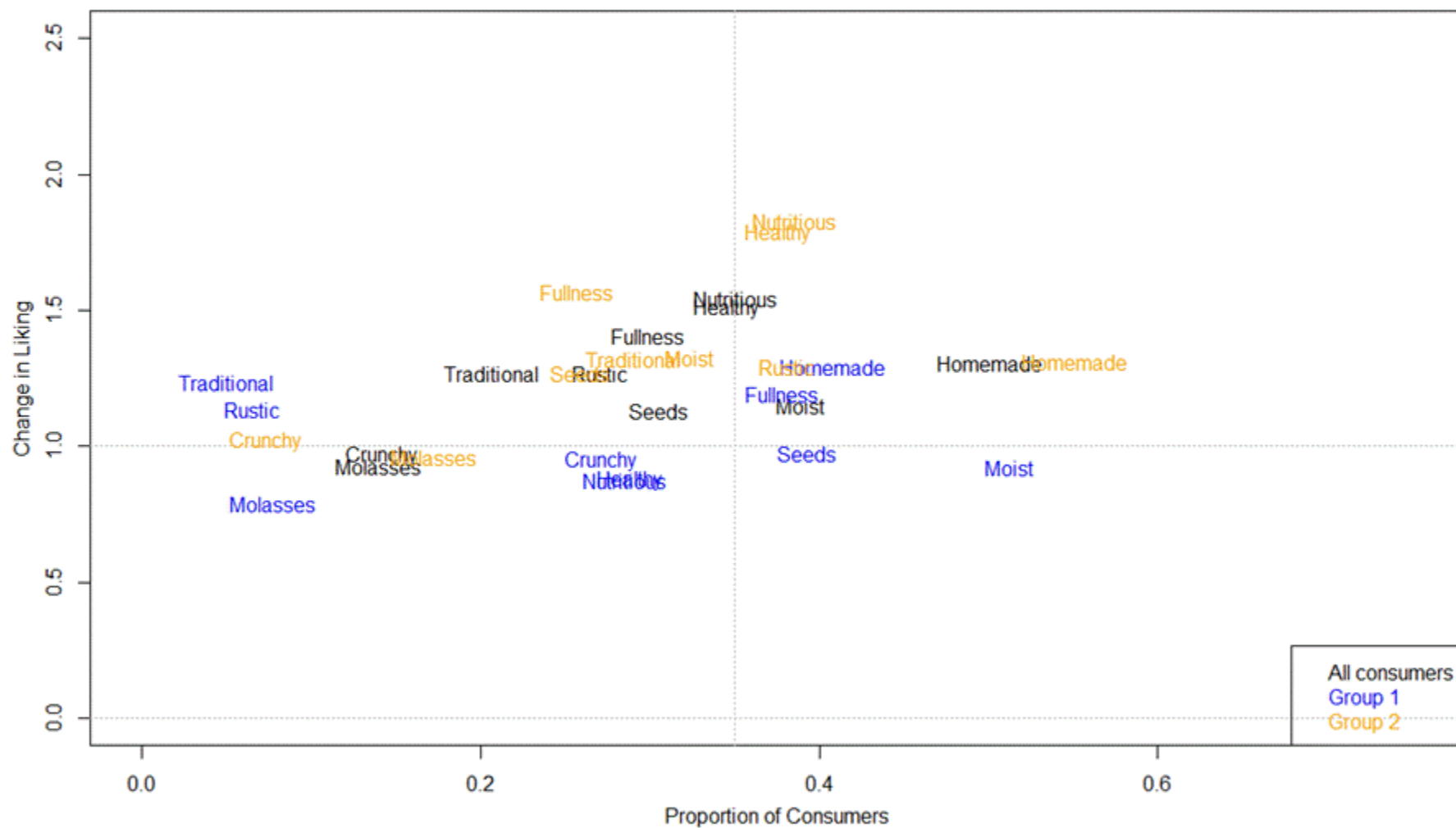
Early results from research (in progress)

# 'Must have' attributes (ideal selected)

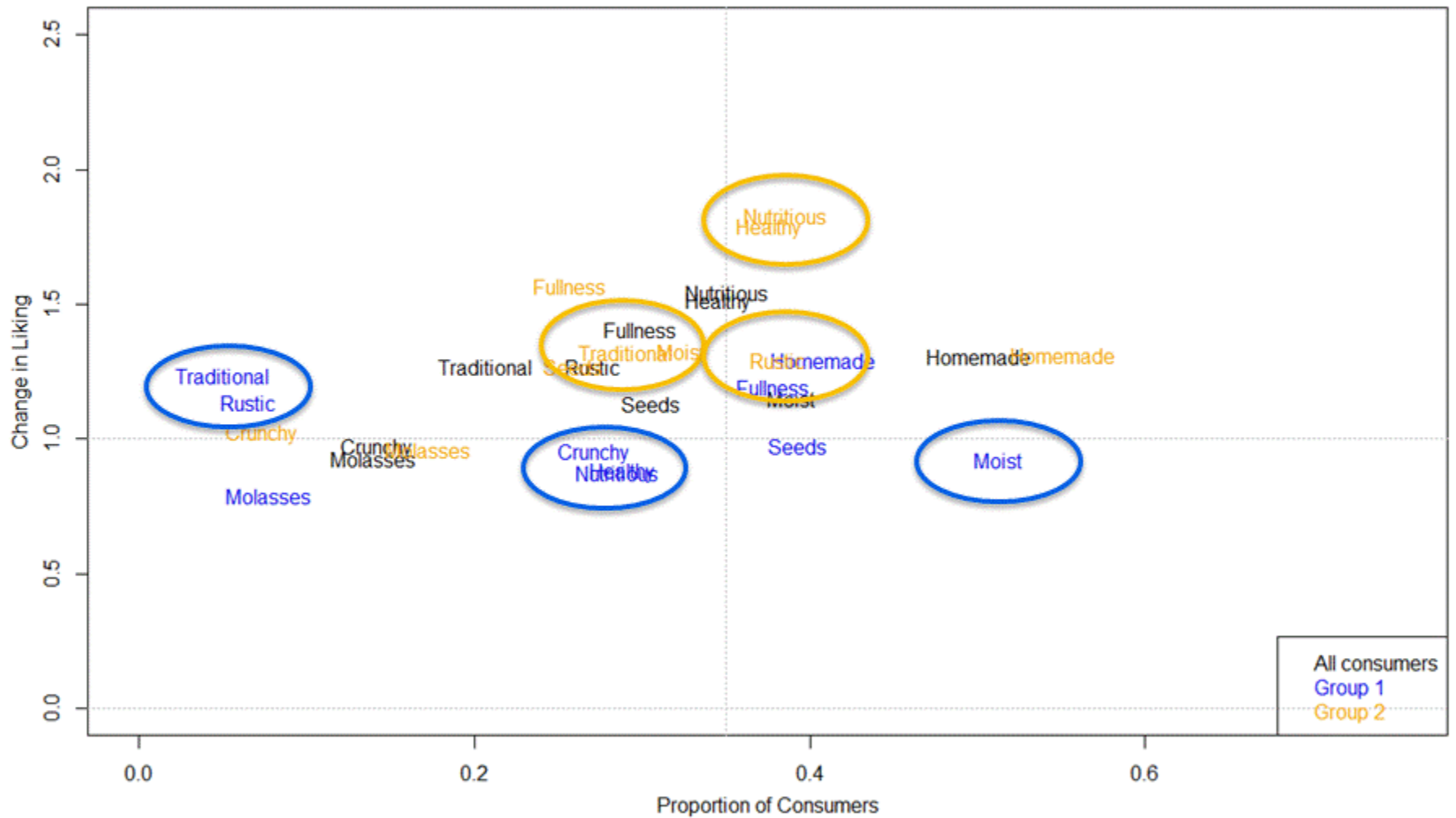




# 'Must have' attributes (ideal selected)



# 'Must have' attributes (ideal selected)



**Check and re-check words** to track changes in the cereal.



0:00

☐

Oat flavour

☐

Corn/corn meal flavour

☐

Chocolate flavour

☐

Bitter taste

☐

Sweet taste

☐

Peanut butter flavour

☐

Crunchy texture

☐

Other

# Temporal Check-all-that-apply (TCATA)

### Sample 527 (Sip 2)

After the prompt to swallow, track changes over time by checking (and re-checking) the attributes below.



0:10

Green

Earthy

Dark Fruit

Heat

Red Fruit

Bitter

Sour

Astringency

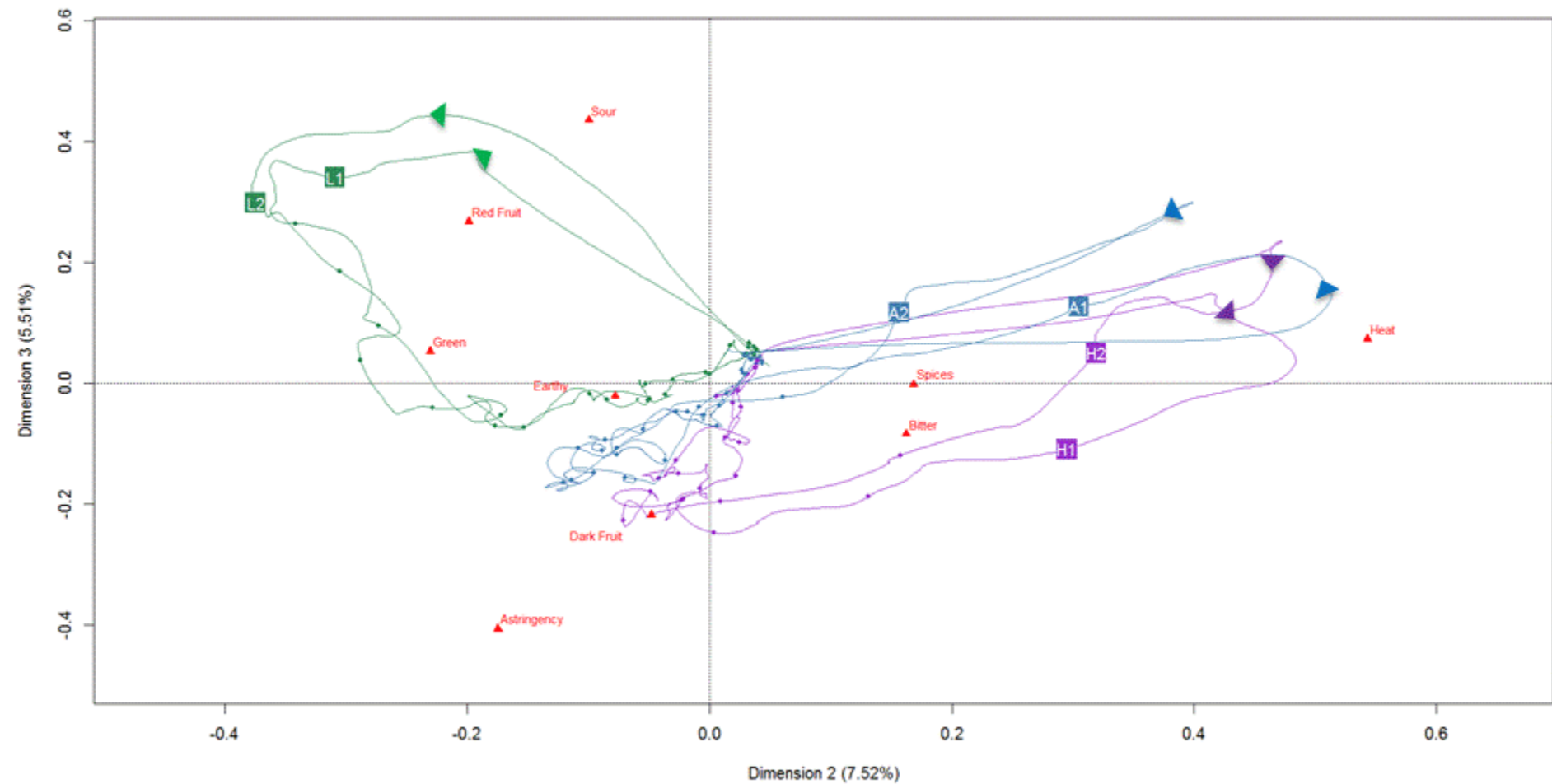
Spice

Other

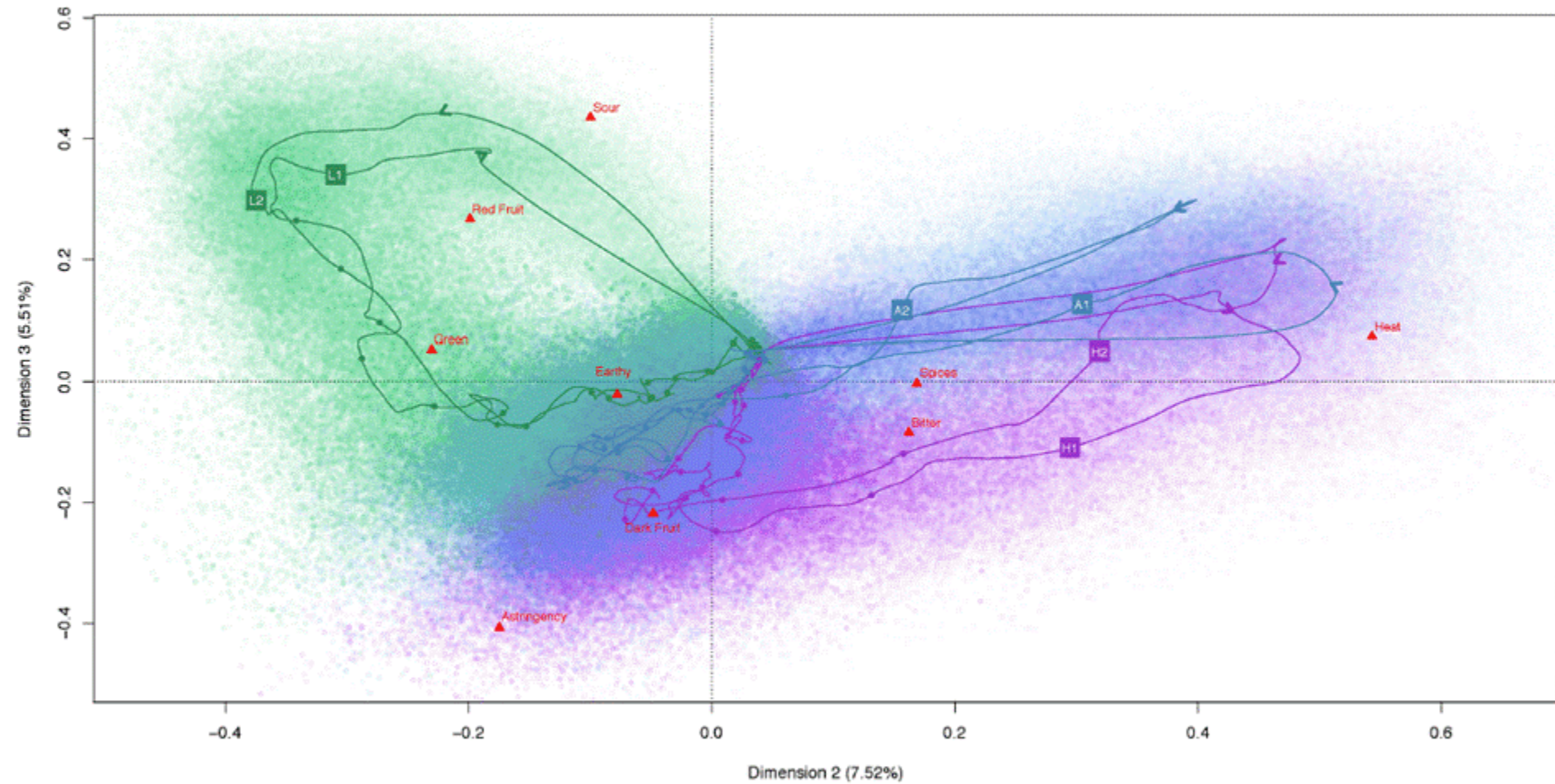
Progress bar with 30 small circular markers, mostly light blue with a few dark blue ones.



# Trajectories



# Trajectories & contrails



# TCATA studies: Question order

TCATA → Liking

Investigate

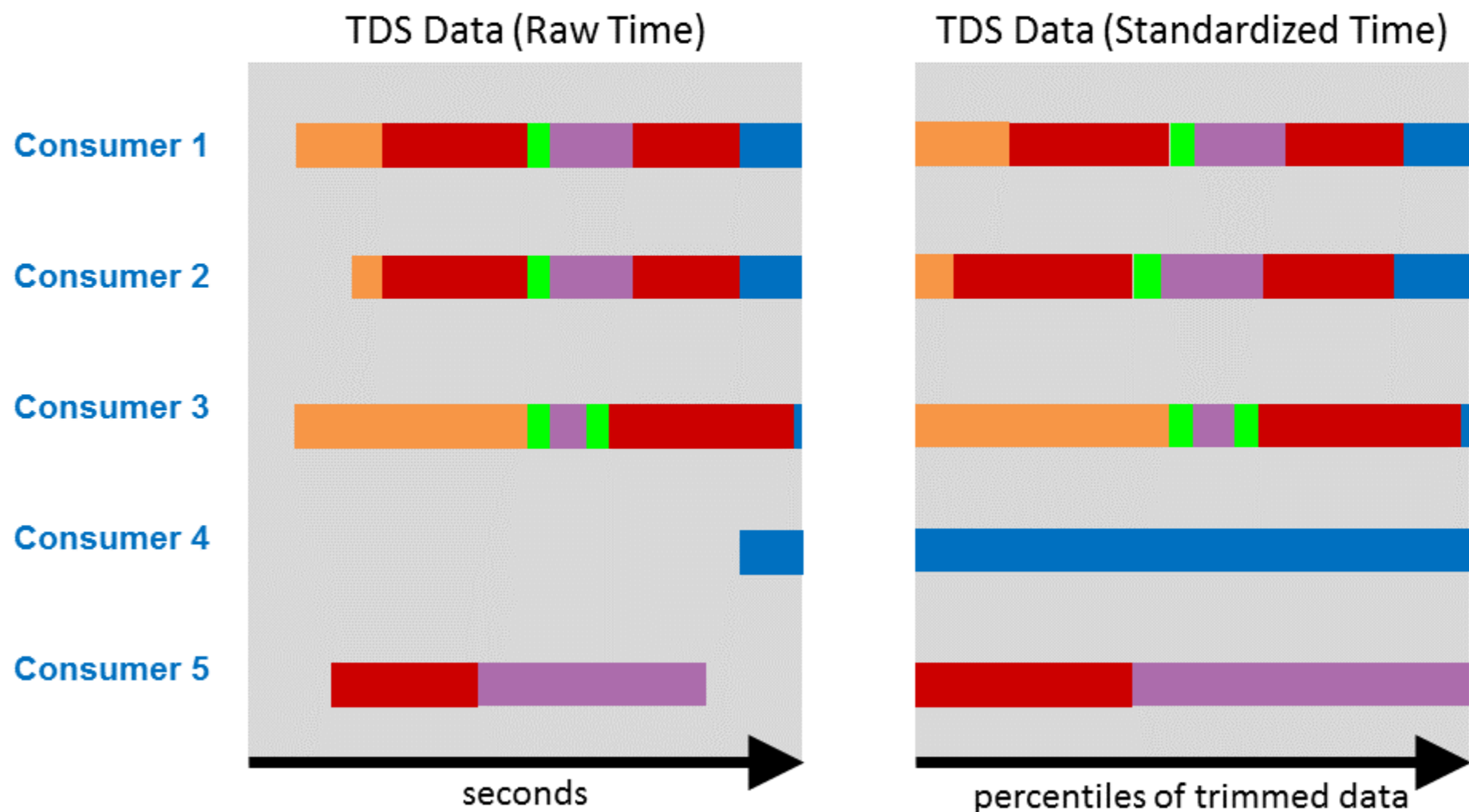
**perception responses** within  
**liking clusters**

and / or

**liking responses** within  
**perception clusters**



# Time standardization



# Time standardization

This can have a dramatic effect on results!

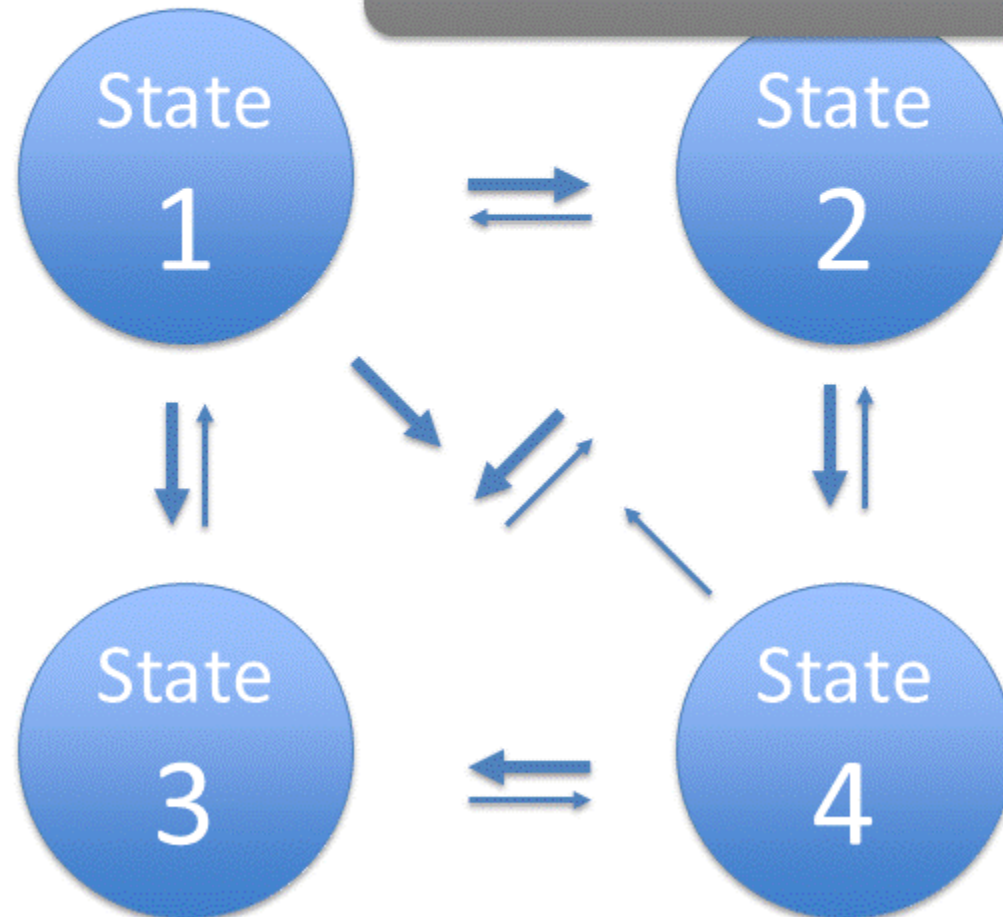
Are we **aligning** or **distorting** the data?

In TCATA evaluations of sparkling wines, **duration of perception** was found to increase with carbonation level... thus **time standardizing removes real product effects!**

**Apply with caution!**

Hidden Markov Models for clustering consumers based on dynamic (TCATA) perception data.

Research in Progress...



# **Part III:**

# **Conclusion**



**“A foolish consistency is  
the hobgoblin of little  
minds...”**

**Ralph Waldo Emerson  
1841**





Does a study that is designed and analyzed in a manner that is consistent with standard practices always make sense?

Of course not.

It's important to follow the design and analysis rules that need to be followed and break the rules that need to be broken.

Which rules are which?

**“Experience is knowing  
when to put your hand in  
the wood chipper.”**

**Chris Findlay**, as quoted by **John Hayes**  
at the Society of Sensory Professionals 2014 Conference  
in Tucson, Arizona

# Selected References

- Browne, R.P. et al. (2013). A partial EM algorithm for clustering white breads. arXiv:1302.6625.
- Castura, J.C., et al. (2016). Using contrails and animated sequences to visualize uncertainty in dynamic sensory profiles obtained from temporal check-all-that-apply (TCATA) data. *Food Quality and Preference* **54**, 90-100.
- Franczak, B.C., et al. (2015). Product selection for liking studies: The sensory informed design. *Food Quality and Preference* **44**, 36–43.
- Hottenstein, A.W., et al. (2008). Preference segments: A deeper understanding of consumer acceptance or a serving order effect? *Food Quality and Preference* **19**, 711–718.
- Lawless, H.T., & Heymann, H. (2010). *Sensory Evaluation of Food*, Food Science Text Series, DOI 10.1007/978-1-4419-6488-5\_9, Springer Science+Business Media, LLC.
- Li, M. (2014). *Model-Based Clustering for Sensory Data and Liking* (Doctoral dissertation). Retrieved from <http://atrium.lib.uoguelph.ca/xmlui/handle/10214/8108>.
- McMahon, K.M., et al. (2017). Perception of carbonation in sparkling wines using descriptive analysis (DA) and temporal check-all-that-apply (TCATA). *Food Quality and Preference* **59**, 14-26.
- McNicholas, P.D. (2017). *Mixture Model-Based Classification*. Boca Raton, FL: CRC Press.
- Meyners, M., et al. (2013). Existing and new approaches for the analysis of CATA data. *Food Quality and Preference* **30**, 309–319.
- Tang, Y., et al. (2014). Model Based Clustering of High-Dimensional Binary Data. *Computational Statistics & Data Analysis* **87**, 84-101. arXiv:1404.3174.
- Wakeling, I., & Macfie, H.J.H. (1995). Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples from t may be tested. *Food Quality and Preference* **6**, 299-308.
- Williams, E.J. (1949). Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Australian Journal of Scientific Research* **2**, 149-168.

# THANK YOU

# MERCI

**John C. Castura**  
VP Innovation & Research

AGROSTAT 2018, édition 2018

