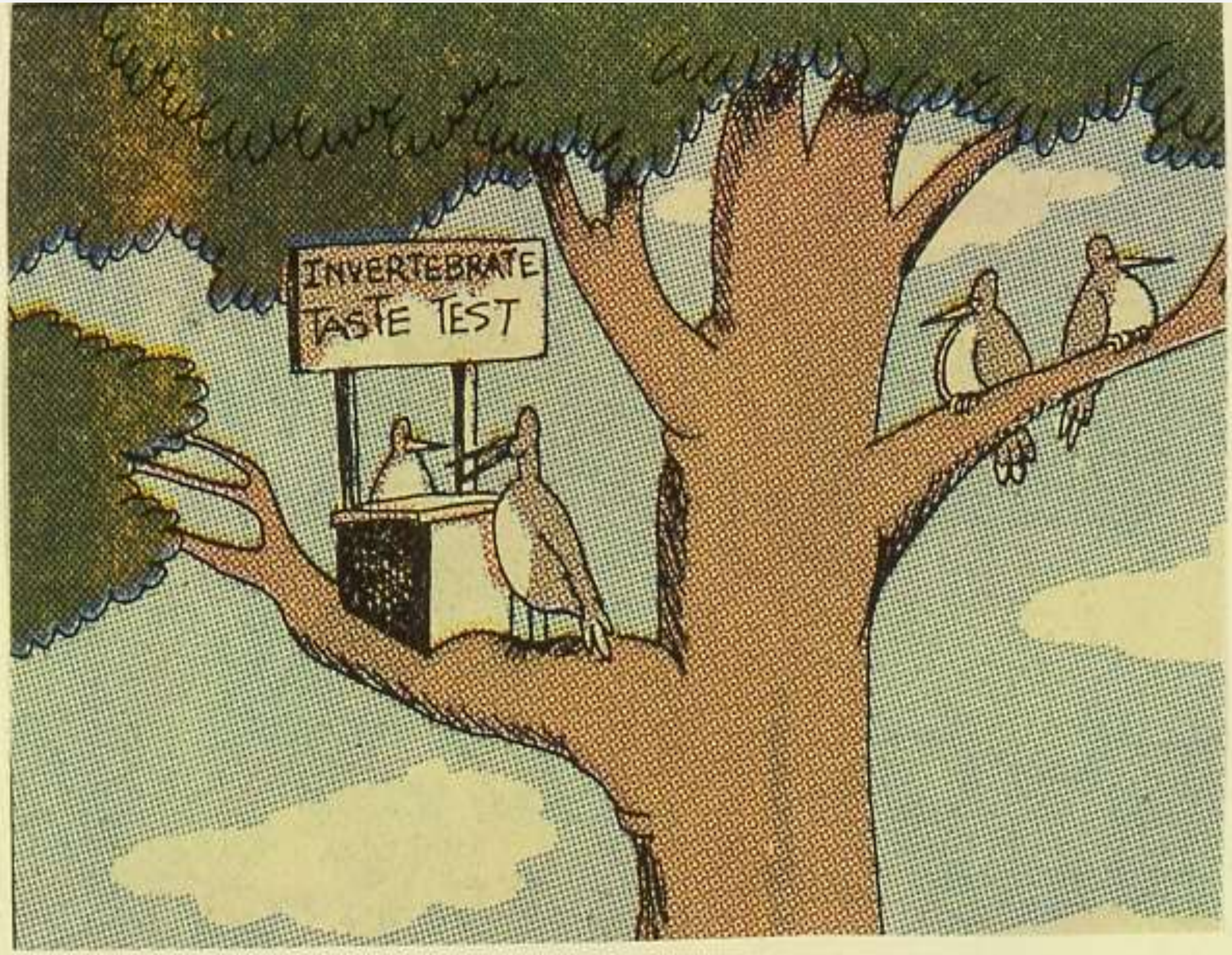
When JAR Scales and Penalty Analysis Tell Different Stories

Harry T. Lawless, Ph.D.
Professor Emeritus
Cornell University
harry.lawless@cornell.edu

Presented at the Regional Meeting, Society of Sensory Professionals
Webcast, 5/22/14



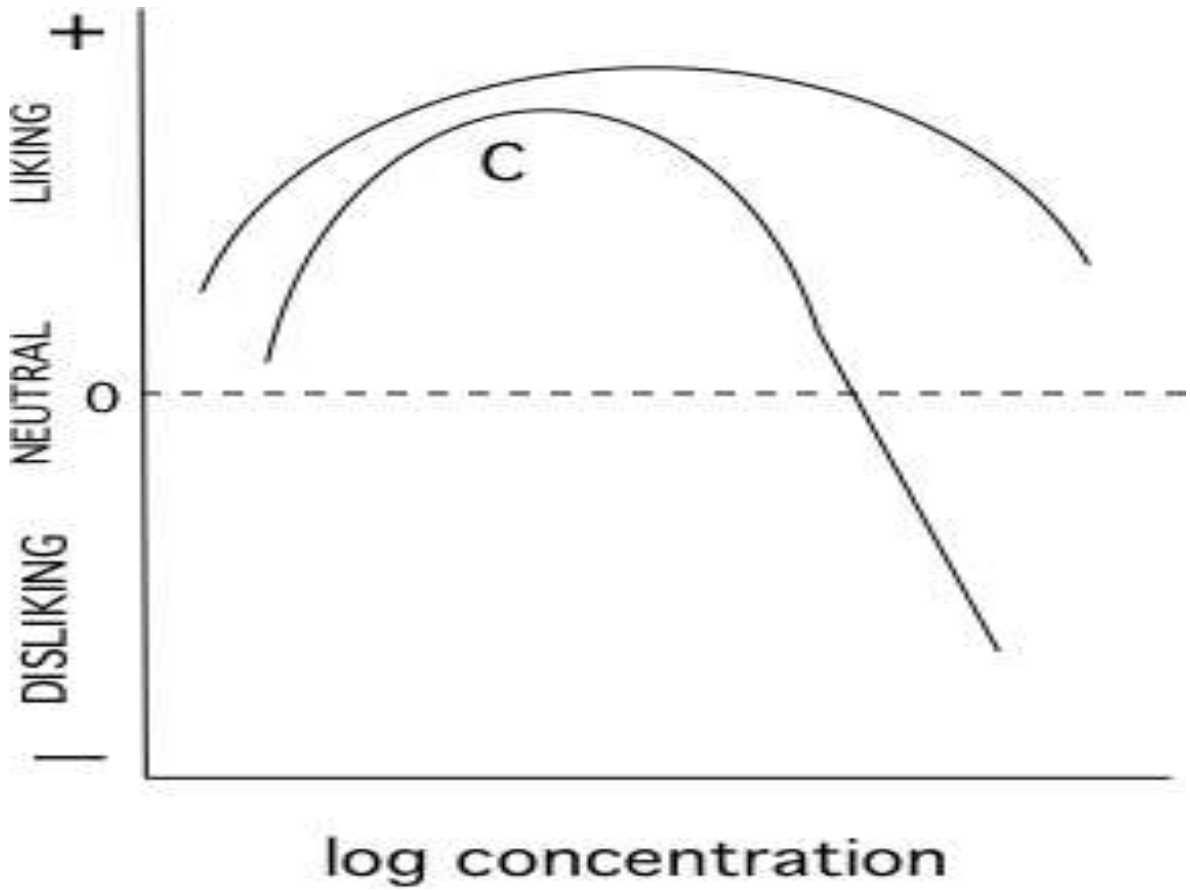
Mmmmmmmm . . . nope . . . nope . . .
I don't like that at all . . .
Too many legs.

“Sweet spot”

- More is better, but only up to a point
- An optimum or “bliss point” may exist
- There can be too much of a good thing. . . .

- But beware individual preferences, segmentation

“Psychohedonic” functions



"….a moderate degree of warmth is pleasant,
and the pleasure increases with the heat to a certain degree,
at which it begins to become painful; and beyond this
the pain increases with the heat, just as the pleasure had done before."
Joseph Priestly, 1775.

Situations

- You have recruited a consumer sample of
  - [loyal / heavy / regular / frequent] USERS of your own successful product.
- You have a modified version of the product
  - E.g. fat reduction, sodium reduction, other nutritional improvement, cost reduction, supplier change, process or packaging change.
- Or you have a new product whose properties you wish to optimize

Examples of Just-about-Right scales

Category:

____	Very much too sweet
____	Too sweet
____	Slightly too sweet
____	Just about right
____	Slightly not sweet enough
____	Not sweet enough
____	Very much not sweet enough

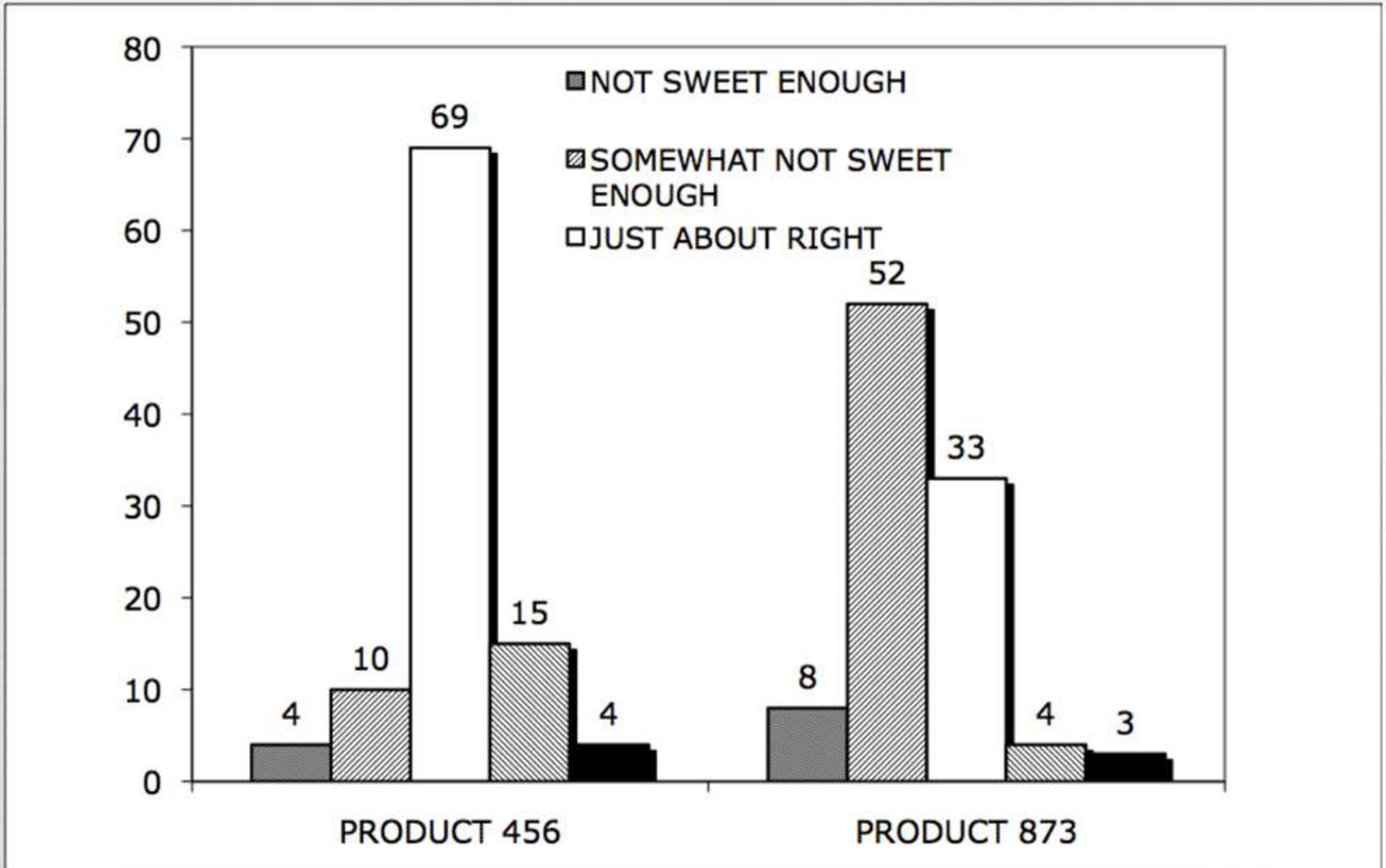
Line Scale:

Not Nearly Sweet enough	Just about Right	Much too Sweet

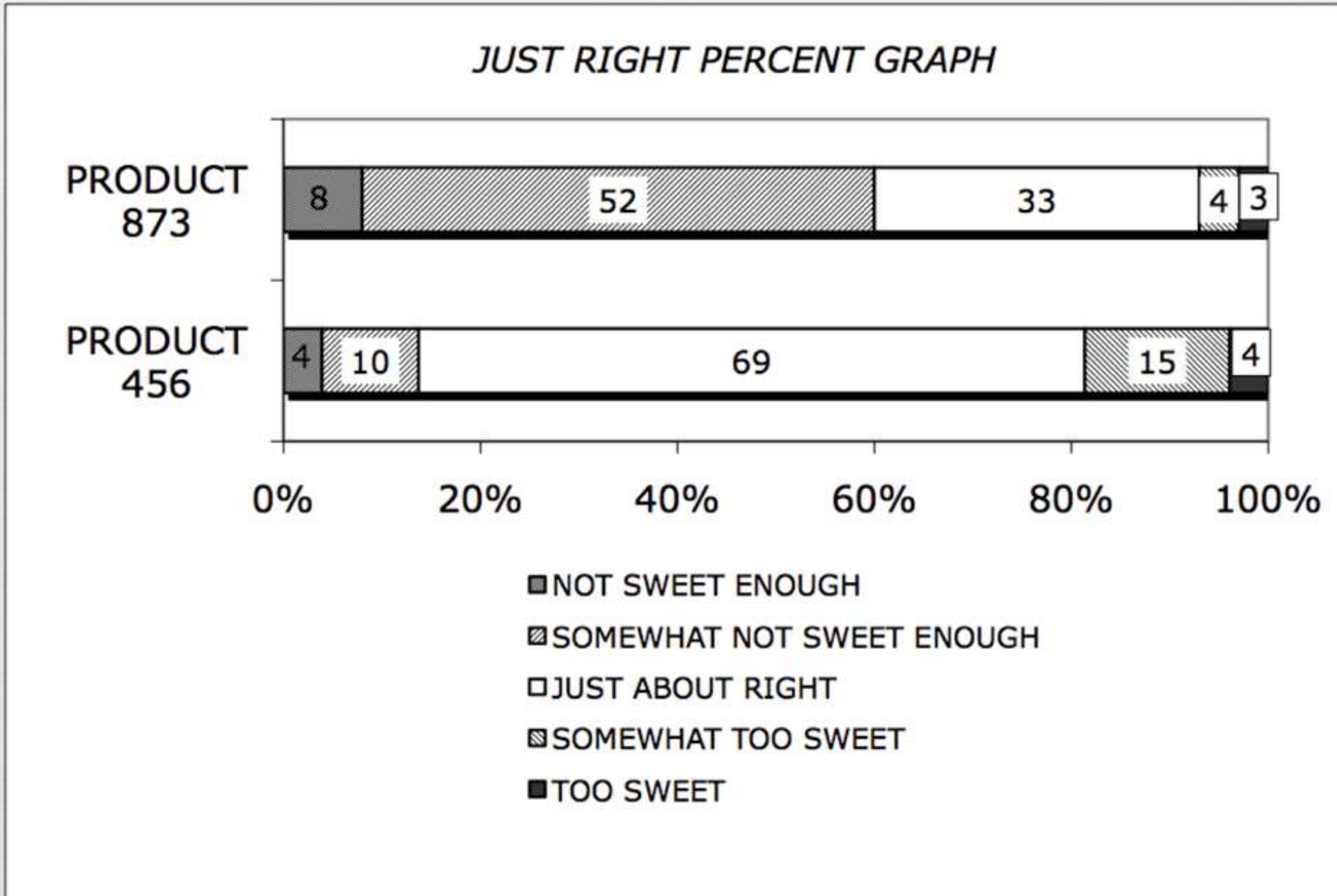
I________________________ I_________________________I

JAR scale desiderata

- 100% of responses on the just-right point is not realistic
- We would like to see
  - A symmetric distribution
  - A peaked distribution (leptokurtic)
  - No more than 20% non-JAR



80
70
60
50
40
30
20
10
0
NOT SWEET ENOUGH
SOMEWHAT NOT SWEET ENOUGH
JUST ABOUT RIGHT
69
52
33
15
10
8
4
4
4
3
PRODUCT 456
PRODUCT 873


JUST RIGHT PERCENT GRAPH
PRODUCT 873
8
52
33
4
3
PRODUCT 456
4
10
69
15
4
0%
20%
40%
60%
80%
100%
NOT SWEET ENOUGH
SOMEWHAT NOT SWEET ENOUGH
JUST ABOUT RIGHT
SOMEWHAT TOO SWEET
TOO SWEET

Suppose your JAR data are skewed?
And there are a lot of non-JAR on one side?

Is this a problem?
Should the product be revised?

Not necessarily!

The important question:

Did it matter?

Was there evidence of a penalty?

Penalty = lowering of scores due to non-optimal characteristic

(i.e. non-JAR)

Penalty Analysis: The big idea

- Penalty analysis combines information
  - From Just-about-right ("JAR") scales
  - And overall liking ratings (9-pt hedonic scale)

Scaled Liking - 9 pt "Quartermaster"

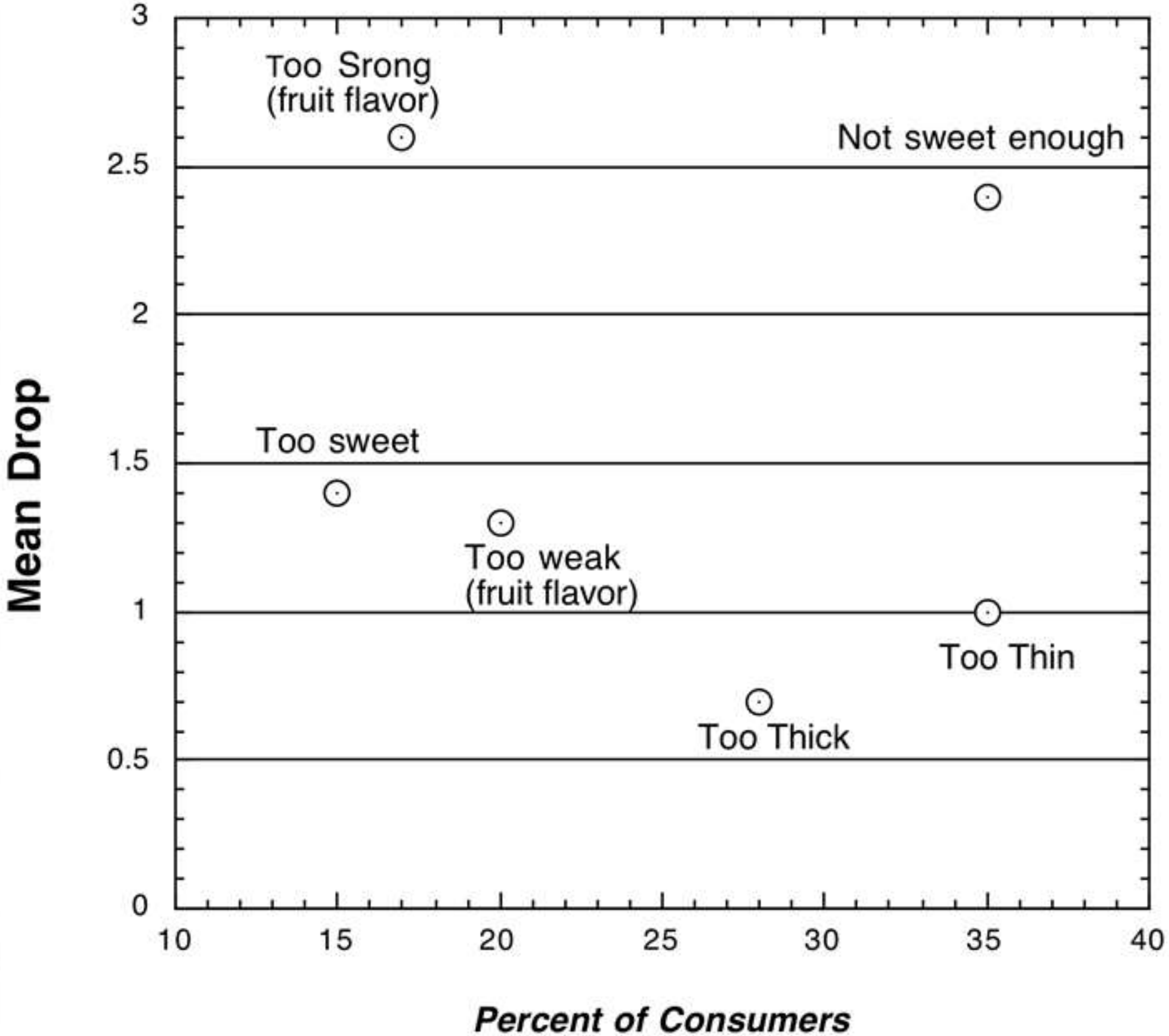
Like extremely
Like very much
Like moderately
Like slightly
Neither like nor dislike
Dislike slightly
Dislike moderately
Dislike very much
Dislike extremely

Adverbs chosen on the basis of psychometric measurement to represent equal spacing (interval scaling)

Penalty defined

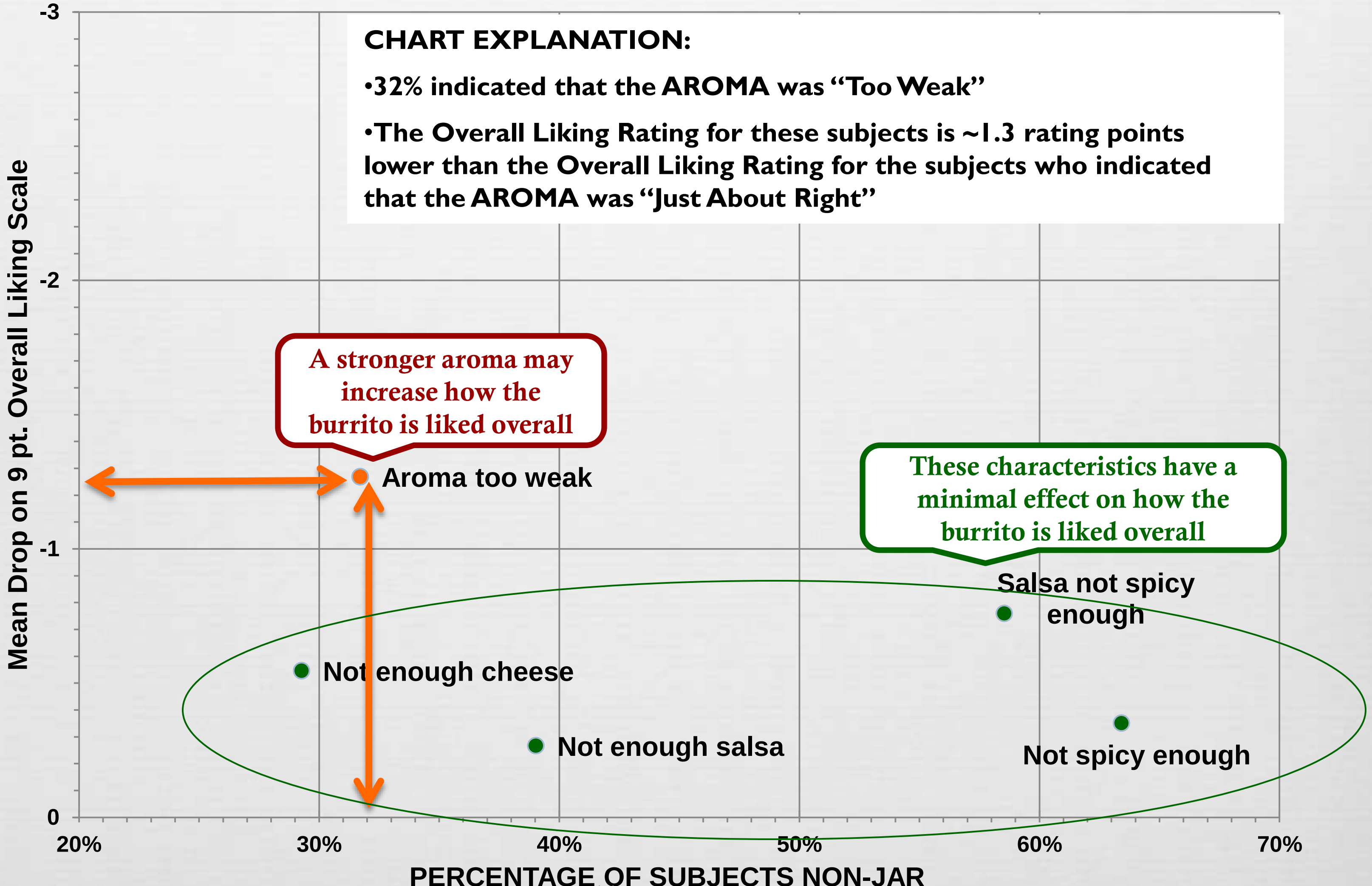
- Penalty = Mean score for persons AT JAR minus
  - Mean score for persons above (or below)
- "Too much" and "too little" calculated <u>separately</u>
  - Produces two "mean drop" penalty values
- Data are usually collapsed to a 3-category analysis
  - Too much, just right, too little.
- Combined with percentage non-JAR value and plotted.

Sample plot



3
2.5
2
1.5
1
0.5
0
10
15
20
25
30
35
40
Mean Drop
Percent of Consumers
Too Srong
(fruit flavor)
Not sweet enough
Too sweet
Too weak
(fruit flavor)
Too Thin
Too Thick

Burrito - New Version

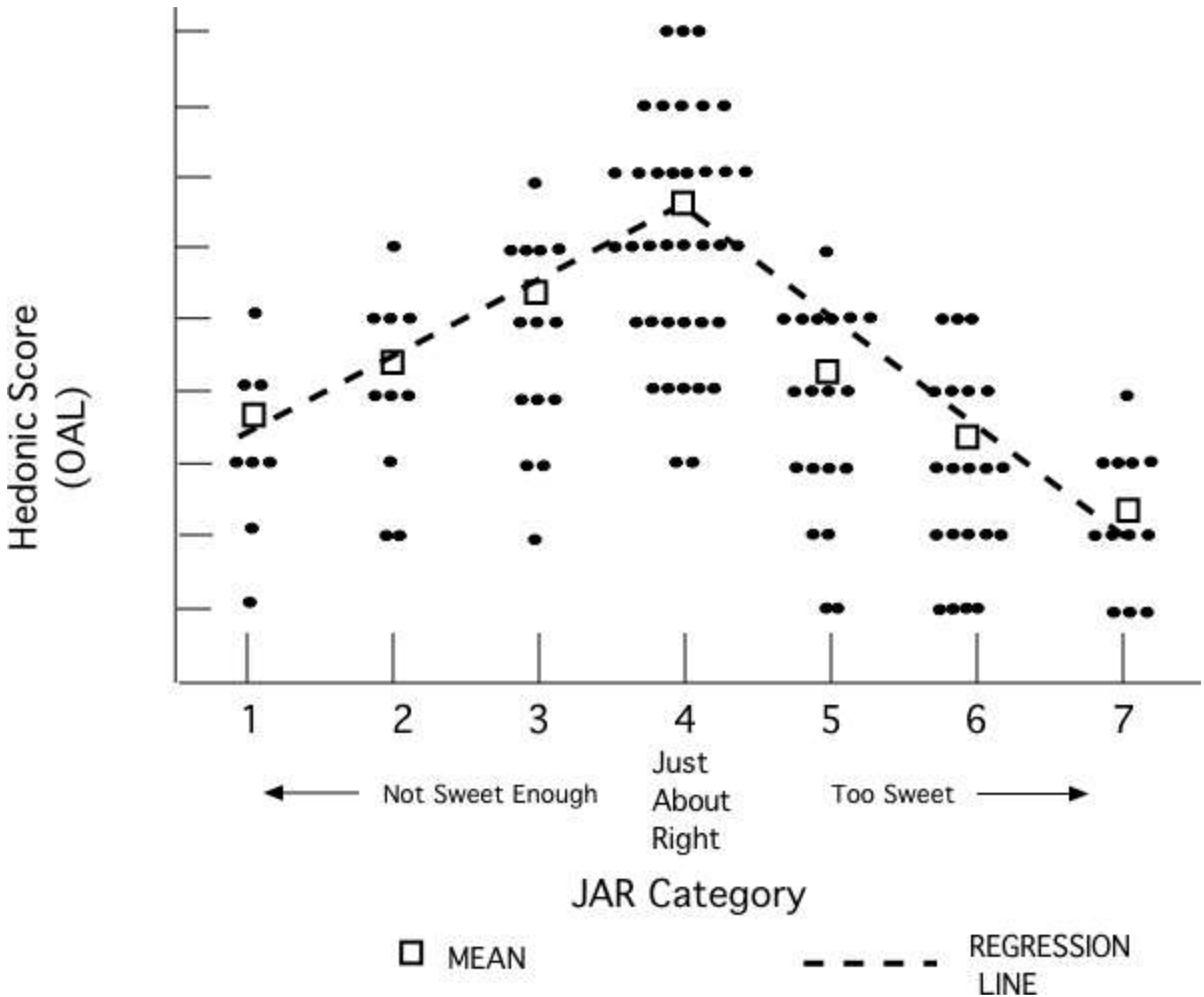


Slope and RSQ*

- Penalty is not just a point!
- Slope can be calculated:
  - How much goodwill do you lose, for each category you move beyond JAR? How fast is the drop-off?
- Also, how tight is that relationship?
  - There is often a lot of spread in the data!
  - R-squared is useful.

*

*



Hedonic Score
(OAL)
1
2
3
4
5
6
7
Just
About
Right
Not Sweet Enough
Too Sweet
JAR Category
MEAN
REGRESSION
LINE

Data handling (each JAR scale)

- Collect JAR data and OAL data in adjacent columns in Excel.
  - Eliminate nonresponders, rows missing data
- Sort (rank) by JAR score
- Create chart (scatter) for scores 5, 4, 3 vs. OAL.
- Fit linear function and get RSQ
- Repeat for scores 3, 2, 1.
- Calculate 3 means for scores of 4 – 5, 3 and 1 – 2.
- Count respondents in those three categories.

Sample data summary

JAR Scale	Measure	Too little	Too much	OAL at JAR	TOTAL (N)
SALTINESS	SLOPE	0.60	0.50		
	RSQ	0.051	0.015		
	MEAN OAL	7.17	7.41	7.91	
	N (%)	47(16%)	23(8%)		292
	DROP	0.74	0.50		

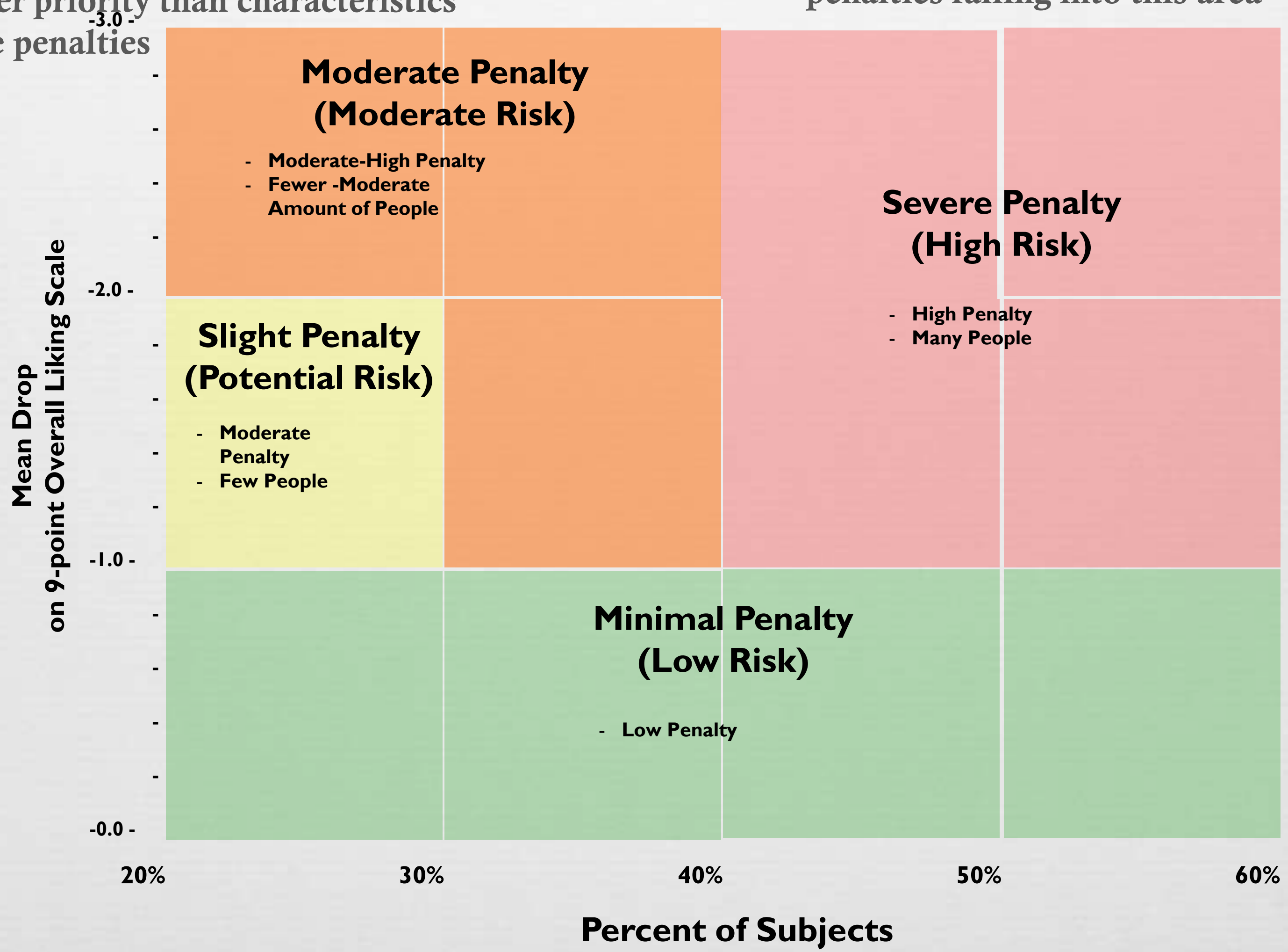
“Zonal Plot”

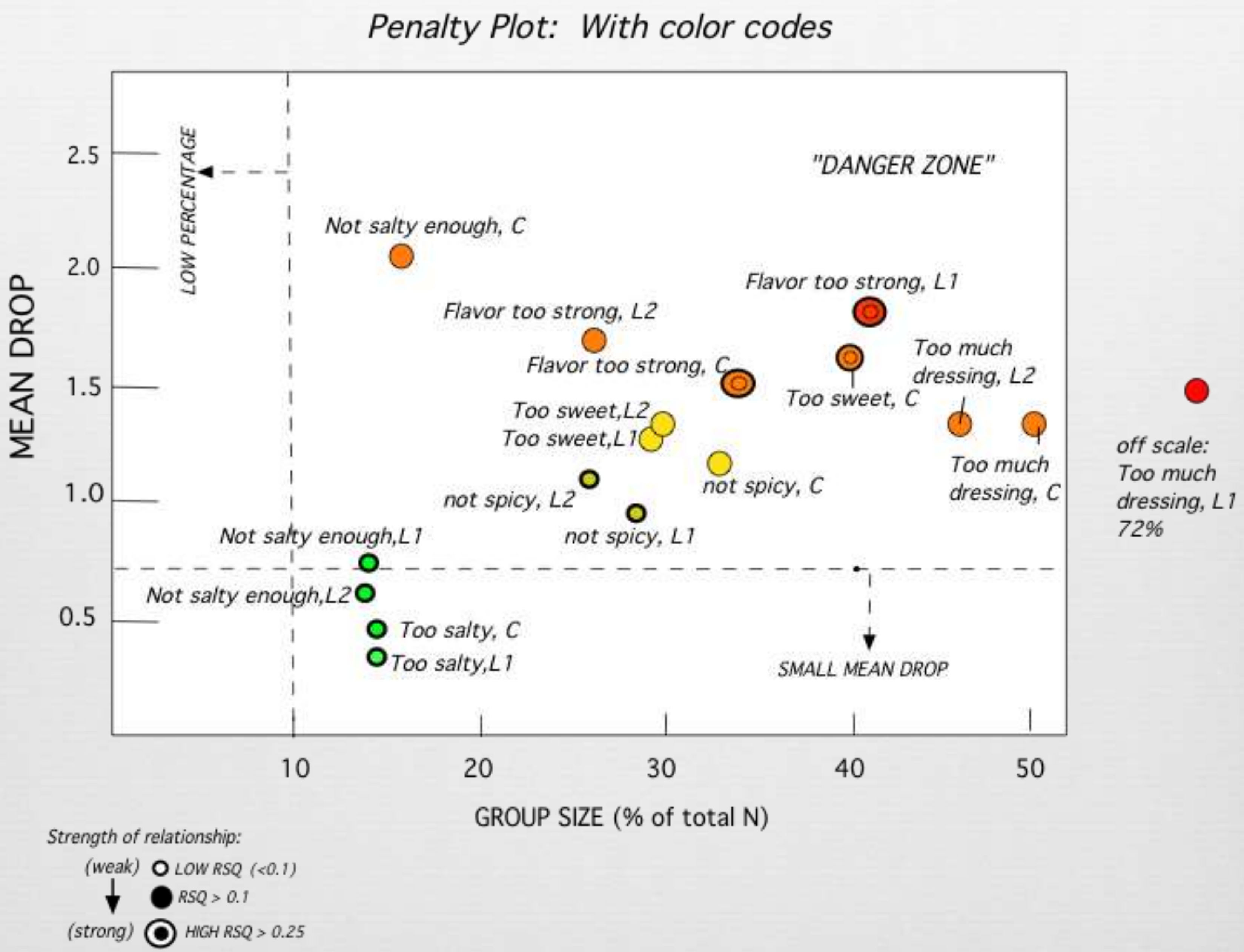
- Can use color coding around action zones
- Parallel to Homeland Security alert colors
  - Green, yellow, orange, red – increasing risk
  - (See next slide)

Zone categories (*from Peryam and Kroll Research*):

Product improvements recommended to address penalties falling into this area, but with lower priority than characteristics having severe penalties

Product improvements recommended to address penalties falling into this area



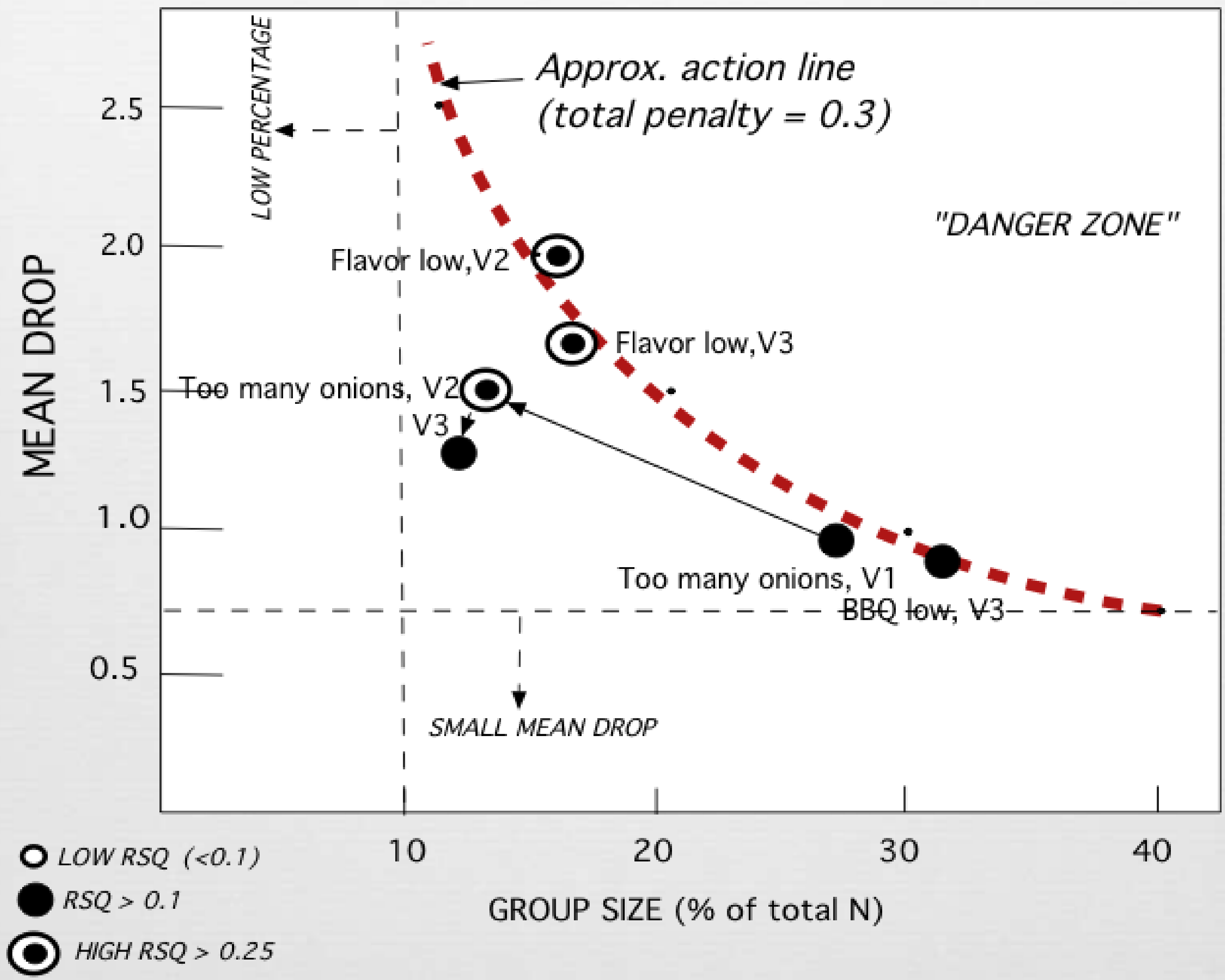


PRODUCT: POTATO SALAD

“Total Penalty”

- Percentage and mean drop multiplied to give a single value
  - E.g. 30% “too sweet” and 1.5 mean drop = 0.45 “total”
  - (A product, not a total)
- Manufacturers can set guidelines or action standards
  - Based on product knowledge
  - History of consumer research and/or complaints
  - E.g. TP greater than 0.30, reject change.
- On penalty plot, forms a curve (hyperbola)

PENALTY PLOT With Total Penalty "action line"



PRODUCT: BBQ BURGER WITH ONIONS

Healthy Dining SBIR Project

Anita Jones-Mueller, MPH, President, HEALTHY DINING
Esther Hill, PhD, Director of Research and Grants
Erica Bohm, MS, VP and Director of Strategic Partnerships
Anjali Patel, MPH, RD, Manager of Research
Nancy Snyder, MS, Graduate Research Assistant
Jonathan Moore, MPH (candidate), Data Coordinator
Sensory Evaluation Consultant: Harry Lawless, PhD
Statistics Consultant: Valentine Njike, MD, MPH

OVERVIEW OF THE RESEARCH PROJECT

- Research supported by the National Cancer Institute of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
- Seeks to reduce levels of calories, saturated fat and/or sodium
- Goal is reduction in target ingredients without negative impact on perceived taste, value or preference among guests
- Further aims:
    - Calculate potential cost savings
    - Quantify potential public health benefits
    - Gather feedback from restaurants

- Confidentiality of restaurant participants and names of their menu items

RESEARCH DESIGN

- Eight total brands, four menu items per restaurant
- For each item, one or more “target” ingredients to reduce:
  - Calories
  - Saturated Fat
  - Sodium
- Three versions tested:
  - Current
  - Level 1 modification (slight reduction, e.g. 10%)
  - Level 2 (moderate reduction, e.g. 25%)
- Four restaurant locations, per brand participant.

RESEARCH METHODS: RECRUITMENT & tasting process

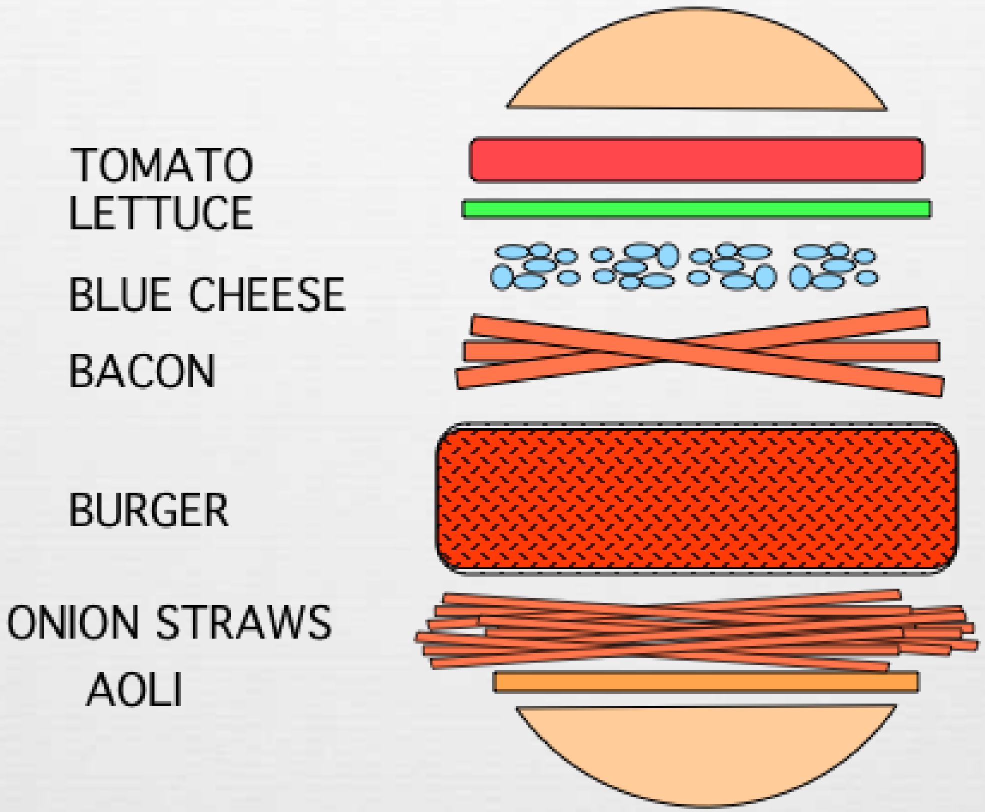
- Participants recruited using corporate customer database and email blasts
- Gift card incentive of $25
- Participants screened for:
  no allergies, customer of the chain/franchise (once every 2-3 months or more), frequency of purchase of test menu items, likelihood to purchase test menu items, age (18 - 70)
- 20-26 customers per taste test session
- Each customer tasted four different menu items, but only one version of each item (randomized of course)
- Scores later compared between current, level 1 and level 2 modification versions

Taste Test Questionnaire – Key measures

- "Overall" variables (Opinion, Appearance, Aroma, Flavor, Texture): 1 to 9 scale
- Open-ended questions: likes and dislikes
- JAR variables (Flavor strength, Amount of "Ingredient," Saltiness): 5 point scales
- Use frequency
- Likelihood to Purchase
- Demographic questions: gender, age, ethnicity and race

Sample product:

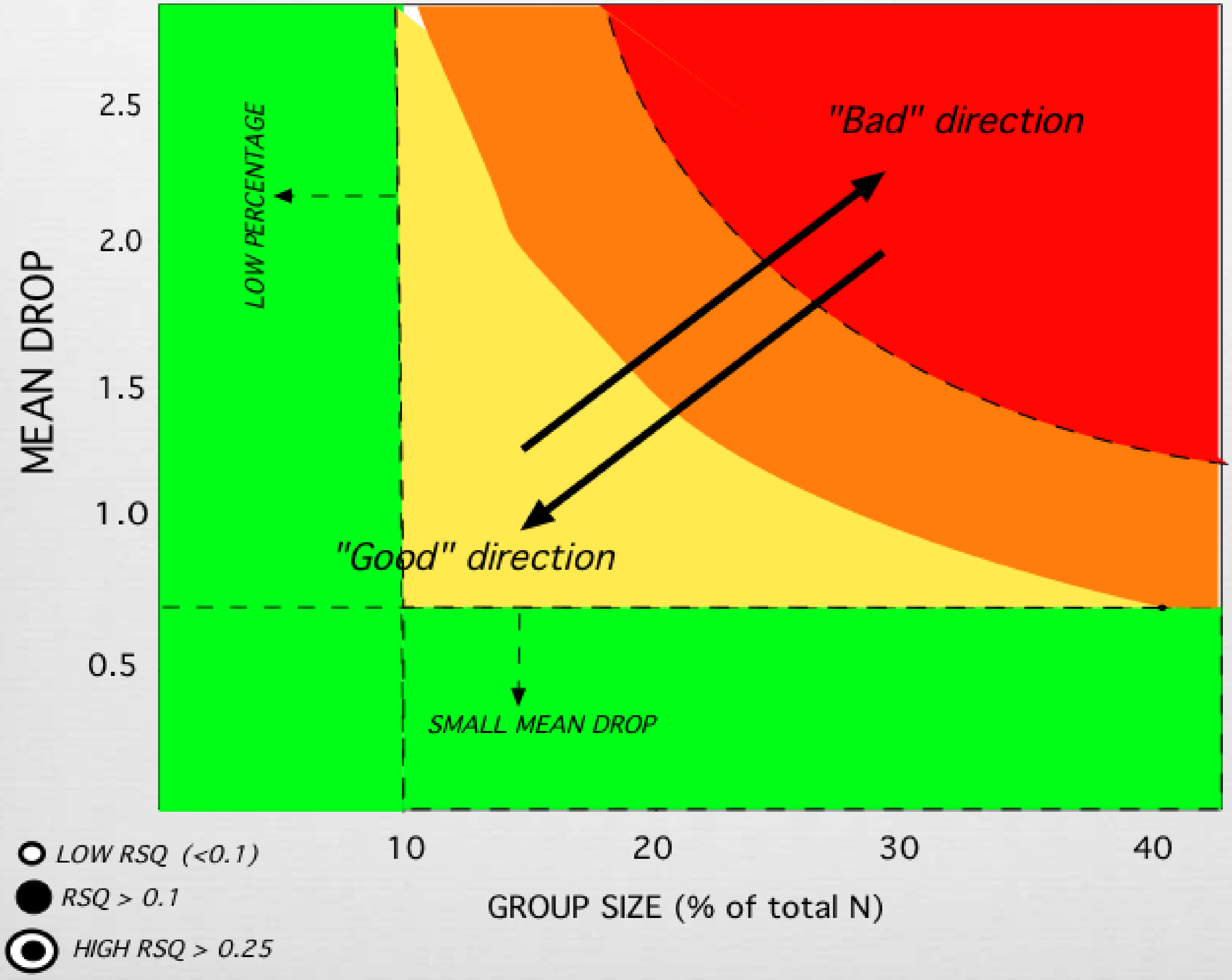
Hypothetical gourmet burger



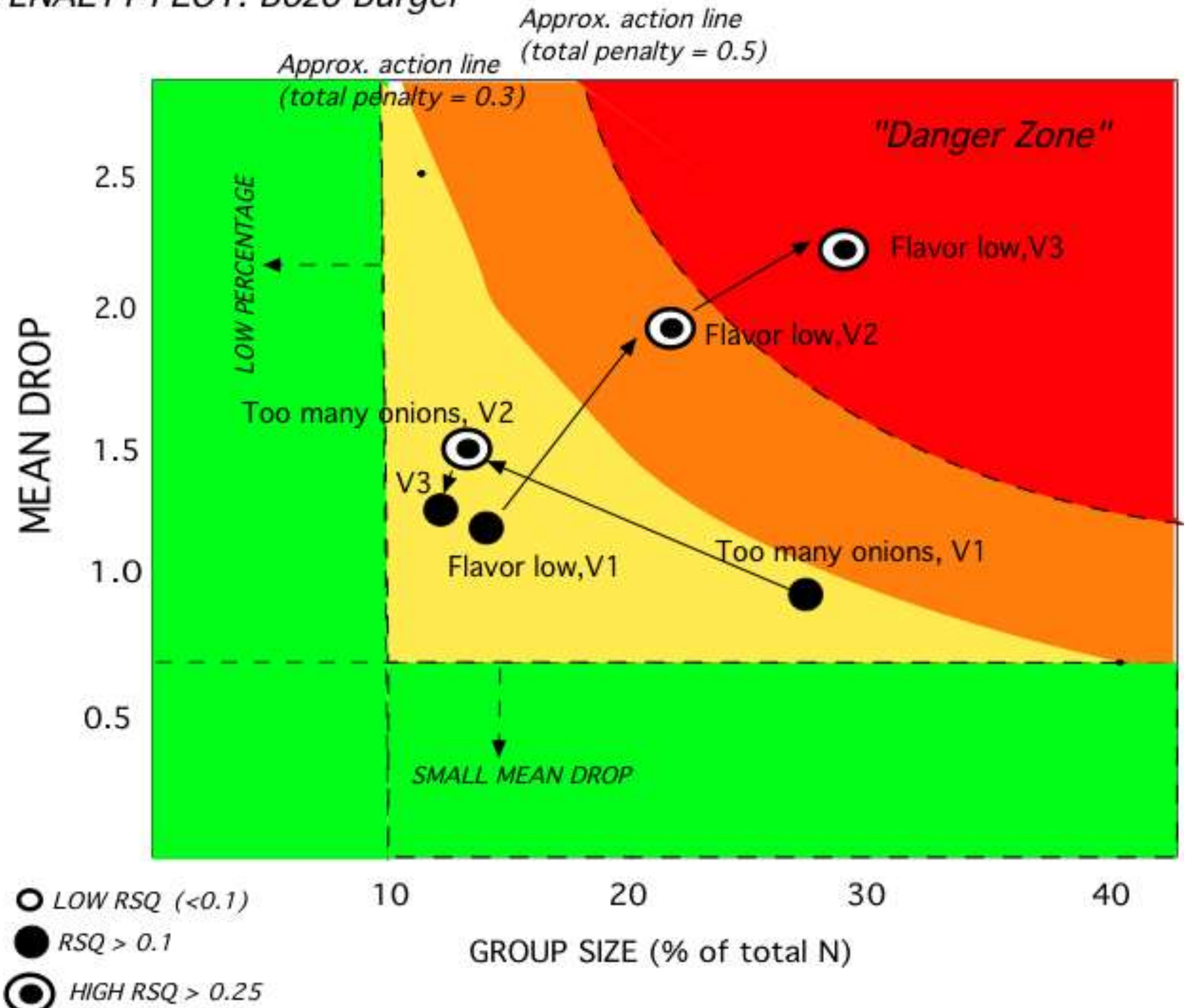
Spike's burger, last week: (bottom to top) cole slaw, fried green tomato, ½ lb burger patty, Pulled pork BBQ, bacon, fried egg. The bacon is optional.

Other products are more difficult to modify, unless portion size is reduced.
Lowering cheese on nachos→ may need to reduce chips (cheese/chip ratio)

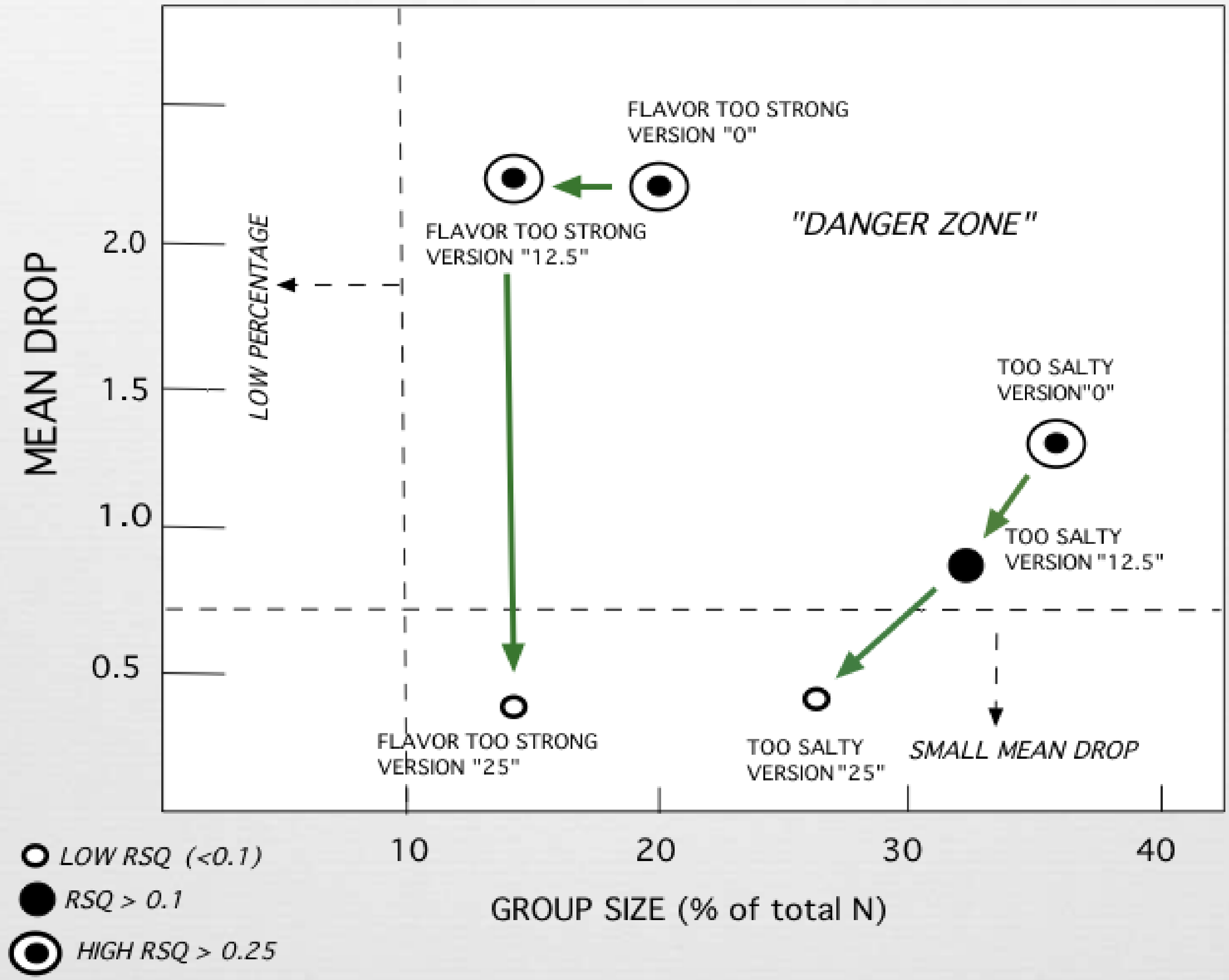
Effects of changing formulations: which way does it move?



PENALTY PLOT: Bozo Burger



PENALTY PLOT: Seafood medley in broth

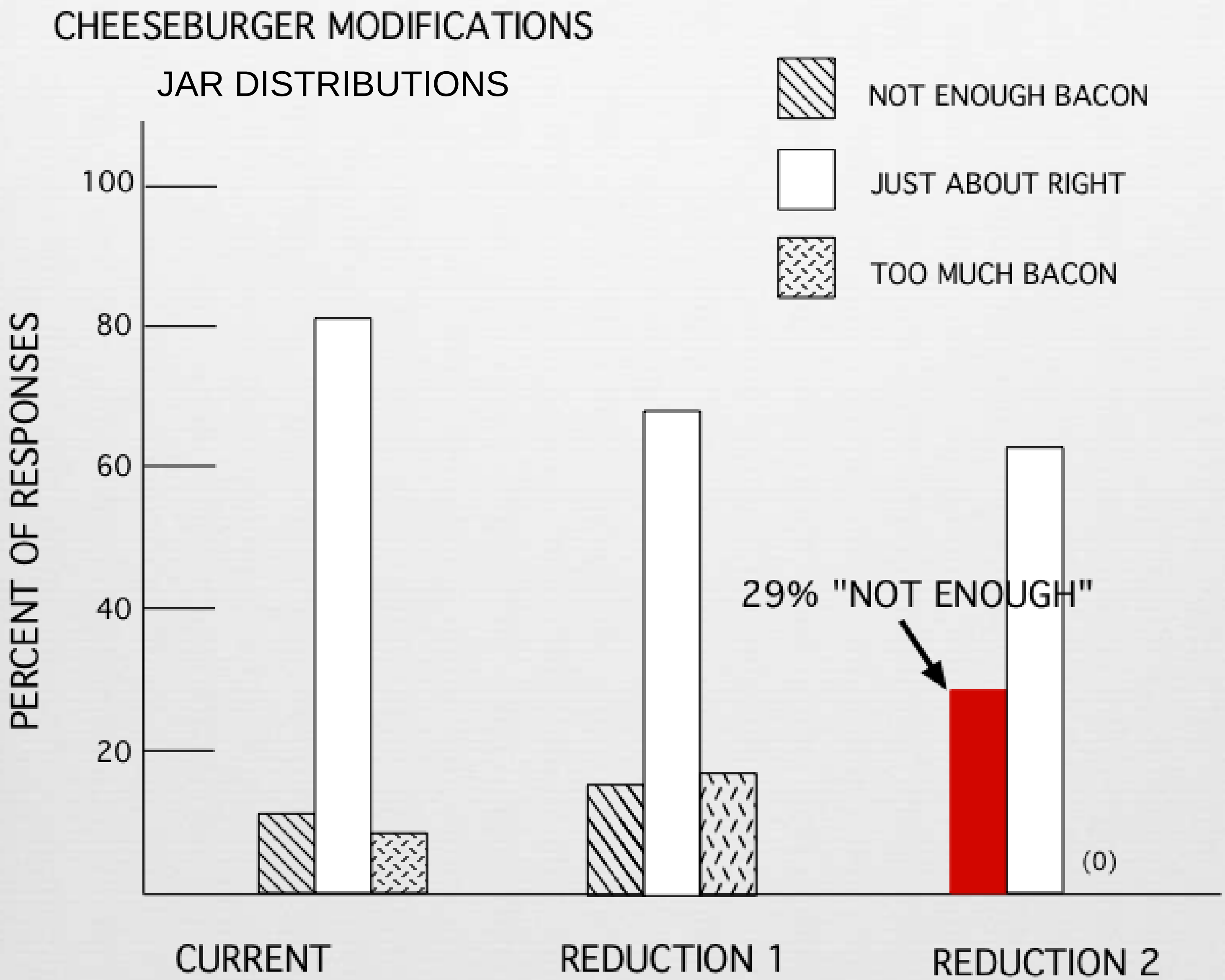


Sodium reduction appears to be a good thing in the new versions.

The ambiguous corners

- High penalty+ large group: Action is clear.
- Low penalty+ small group: Action not needed.
- High penalty +small group: What to do?
  - There aren't a lot of them, but they are dissatisfied.
- Low penalty + large group: Can I ignore it?
  - Maybe this doesn't matter?
  - Everybody would like more bacon……

Note: The yellow corners are zones where JAR info and Penalty info DIFFER!



IS REDUCTION 2 A PROBLEM? RISKY?

Issue: the bacon dilemma

- Low penalty, large group
- With some attributes, most everyone would like more
  - More bacon is simply better!
- But does it move the needle if you change it?
- Low penalty suggests it is “not a big deal”
- Cost-benefit analysis may help
  - What is the cost of adding more bacon?
  - Does it make a difference?

Bacon dilemma, cont.

- This is where JAR analysis and penalty analysis give different conclusions and different courses of action
- JAR shows there is a large non-JAR group.
  - On that basis alone, action would be recommended!
- Penalty analysis shows it may not matter!
- The take-home lesson: JAR analysis on it's own may not be conclusive

Another example: Amount of Feta on Greek salad

"Not enough cheese"	Current salad	Reduction level 1	Reduction level 2
Group Size	34%	42%	39%
Mean drop	0.48	0.25	0.02

Conclusion: Although JAR indicates a potential problem due to large group sizes for NON-JAR responses, penalty suggests a reduction in cheese content is LOW RISK.

Another corner: salt

- Issue: Small group but large penalty
  - (upper left corner)
- Halophobic, do not like salt
- Too salty ratings infrequent but strong correlation with mean drop in OAL
- What to do?

Issue: opposite opinions

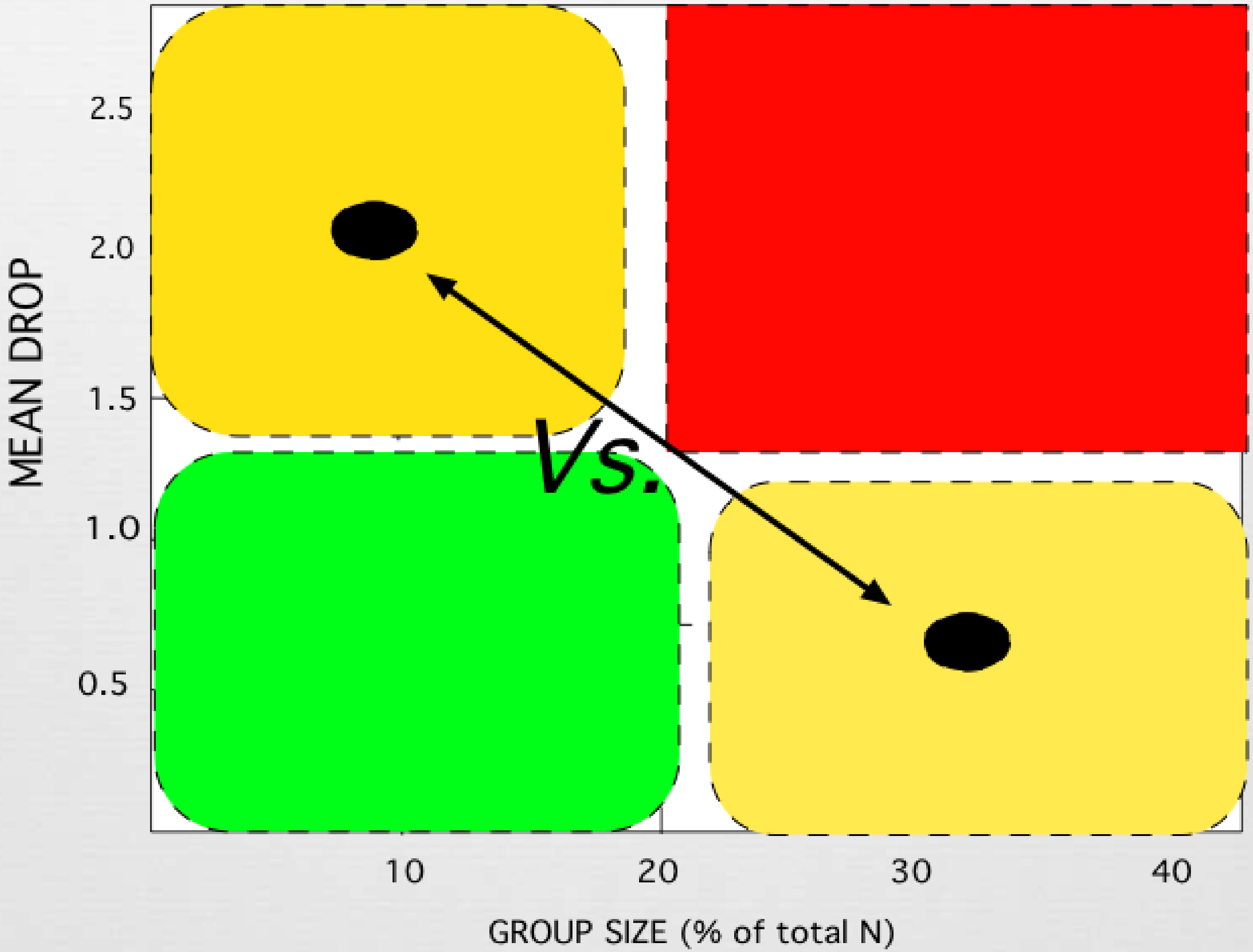
examples: spicy heat level, blue cheese
"polarizing" ingredients/attributes/flavors

- Two NON-JAR groups for the same attribute:
  - one indicating too little
  - and one indicating too much.
- What action?
- Both Low penalty: possibly ignore it.
- Both high penalty: maybe you need two versions . . .
  - New products? Meets a consumer need?
  - Evidence of market segmentation

When they occupy the same part of the penalty space, there is no differentiation i.e. they must be considered equally valid, equally actionable…..

Same corner, Opposite opinion

- Can occur with “polarizing” ingredients
  - E.g. blue cheese on a burger
  - Some people want a lot, some just a smidge.
  - “love it or hate it?”
- Consider two versions of the product?
- Which group has the higher frequency purchaser?
  - Is there an 80/20 rule?



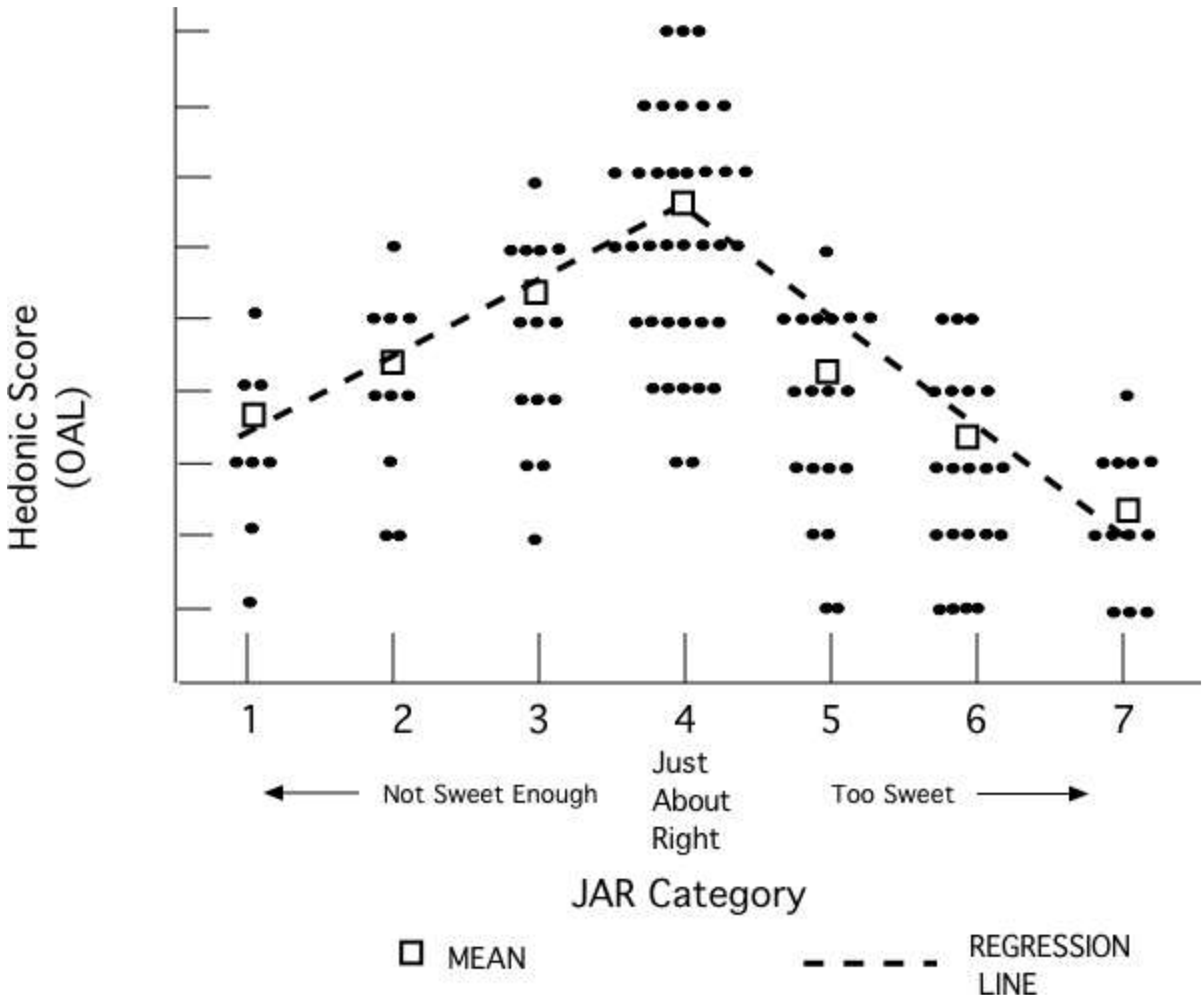
MEAN DROP
2.5
2.0
1.5
1.0
0.5
10
20
30
40
GROUP SIZE (% of total N)
Vs.

Issue: Opposite corners

- Two groups:
  - One large, low penalty,
  - One smaller, high penalty
- Example: Cheese soup, spicy JAR scale
  - Too little (not spicy enough): 26%, mean drop = 0.59
  - Too spicy: 12%, mean drop = 1.46
- Modifying the product to please one group may annoy or alienate the other group. . . .

Caveat: Leverage

- A few respondents at the ends of the JAR scale, with large penalties
- Moves the slope and the RSQ value!
- Can cause an artificial inflation, or be misleading
- Look at the data carefully!
  - Examine the end categories.
  - Could be a "vocal minority" driving the relationship.


Hedonic Score
(OAL)
1
2
3
4
5
6
7
Just
About
Right
Not Sweet Enough
Too Sweet
JAR Category
MEAN
REGRESSION
LINE

Example: salt and oil reduction in hummus

- JAR scale: Flavor strength
- OAL for group at JAR = 8.2 *(not bad!)*
- OAL for group, "flavor too strong" = 6.0
  - Mean drop of 2.2 points, *looks like trouble?*
- But group size = 5 people (only 5% of test group)
  - Only ***one person*** scored below 5 on 9-pt scale

Conclusion: Penalty misleading, JAR group size indicates low risk.

Caveat: correlation is not necessarily causal

- Penalty analysis is purely a mathematical exercise
- Just because there is a drop, it may not be due to this JAR attribute!
- Another item, correlated with this JAR attribute could be the real culprit
- E. g. "not sweet enough" could be due to too much acid.

A few reminders . . .

Check RSQ!

- Example: Amount of cheese on cheeseburger
  - 25 % group size for "too little." Actionable?
  - Mean drop = 0.26 (small)
    - So this is a lower right quadrant penalty
  - RSQ = 0.01 (very low = weak relationship)
    - = BIG spread in the data, hedonic scale.
    - Inconsistent pattern
    - Action *may not be needed*

Check open-ended Q's

Verbatim comments should line up with penalty and/or JAR conclusions.

Look at *distributions*, <u>not just the mean scores</u>

- JAR Scales are a great example of where you can get into trouble by looking only at mean scores.
  - You can have a perfectly centered mean score, but people on both sides of the scale who are very unhappy!
- The same principle holds for 9-pt hedonic data.
- We develop "alienation" scores summarizing the bottom end of the distribution for product modifications. ("bottom 4 box score")

For further information and any additional questions:

harry.lawless@cornell.edu

~~Coming soon~~ to a bookstore near you:

Lawless, H. T.

Quantitative Sensory Analysis,

Psychophysics, Models and Intelligent Design

Wiley/Blackwell (2013)