



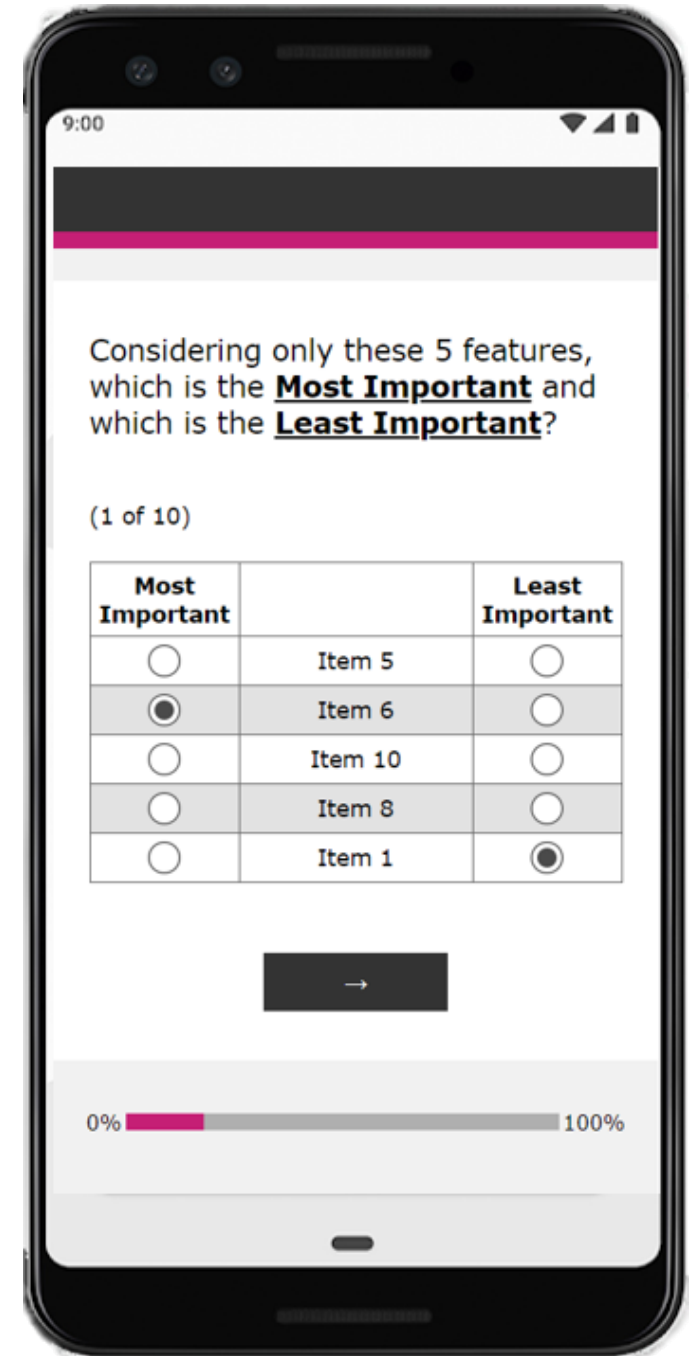
# The Peaks & Pitfalls of MaxDiff at P&G

David Hengehold & Megan Peitz



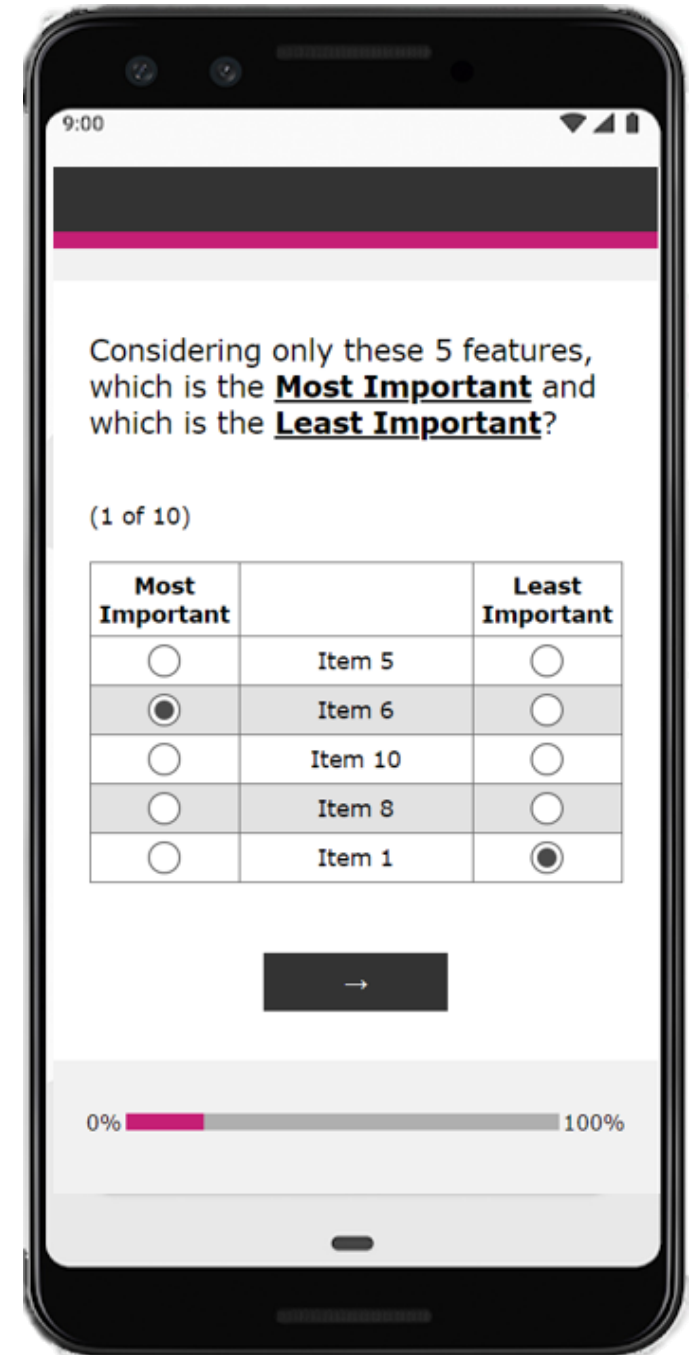
# What is MaxDiff?

...an approach for measuring preferences for a list of items. “Items” can include advertising claims, product benefits, product messaging, images, product names, claims, brands, features, packaging options, and more!



# Quick Review

- Respondents typically shown 2-6 items at a time, asked to indicate which is best and which is worst
- Task is repeated many times, showing a different set of items in each task
- Resulting model provides ratio-scaled scores for each item that can be transformed into ranks



# Why do we like it?

## **MaxDiff Data > Ratings Data**

---

- Greater discrimination among items
- Greater discrimination between respondents on the items
- No scale bias
- Looks great on mobile!

## **MaxDiff Data > Rankings Data**

---

- Ranking is impractical with >10 items
- Ordinal scale results
- What do you report? Top Rank? Top 1+2?







Check out Bryan Orme's article – How Good is Best-Worst Scaling on the Quirks repository - <https://www.quirks.com/articles/how-good-is-best-worst-scaling>

When respondents are  
forced to make **difficult**  
**tradeoffs**,  
we learn what they  
**truly value**

# The details are lost on me...

- Pitfall #1 – Viewing MaxDiff stimuli side-by-side can limit visibility of detailed information.
  - Especially problematic for mobile unless scrolling is used.

- Peak – Panelist previews all thumbnail images which launch (near) full-screen images
  - Then panelists can review them and make best / worst choices.

Option 1	Option 2	Option 3	Option 4
			
Mad Angles	Tedhe Medhe	Tangles	Yumitios
100 gm	125 gm	150 gm	200 gm
Rs.7	Rs.10	Rs.15	Rs.20
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Most Prefer		Least Prefer
<input type="radio"/>		<input type="radio"/>
<input type="radio"/>		<input type="radio"/>
<input type="radio"/>		<input type="radio"/>
<input type="radio"/>		<input type="radio"/>

Preview each product, then select the one you would be Most Interested in purchasing, and which one you would be Least Interested in purchasing. (1 of 17)

## Most Interested

<input type="checkbox"/>	
<input type="checkbox"/>	
<input type="checkbox"/>	
<input type="checkbox"/>	

1  
2  
3  
4

\$1.29/Bar



8 - 1.41 OZ (40g) BARS/  
NET WT 8.46 oz (240g)

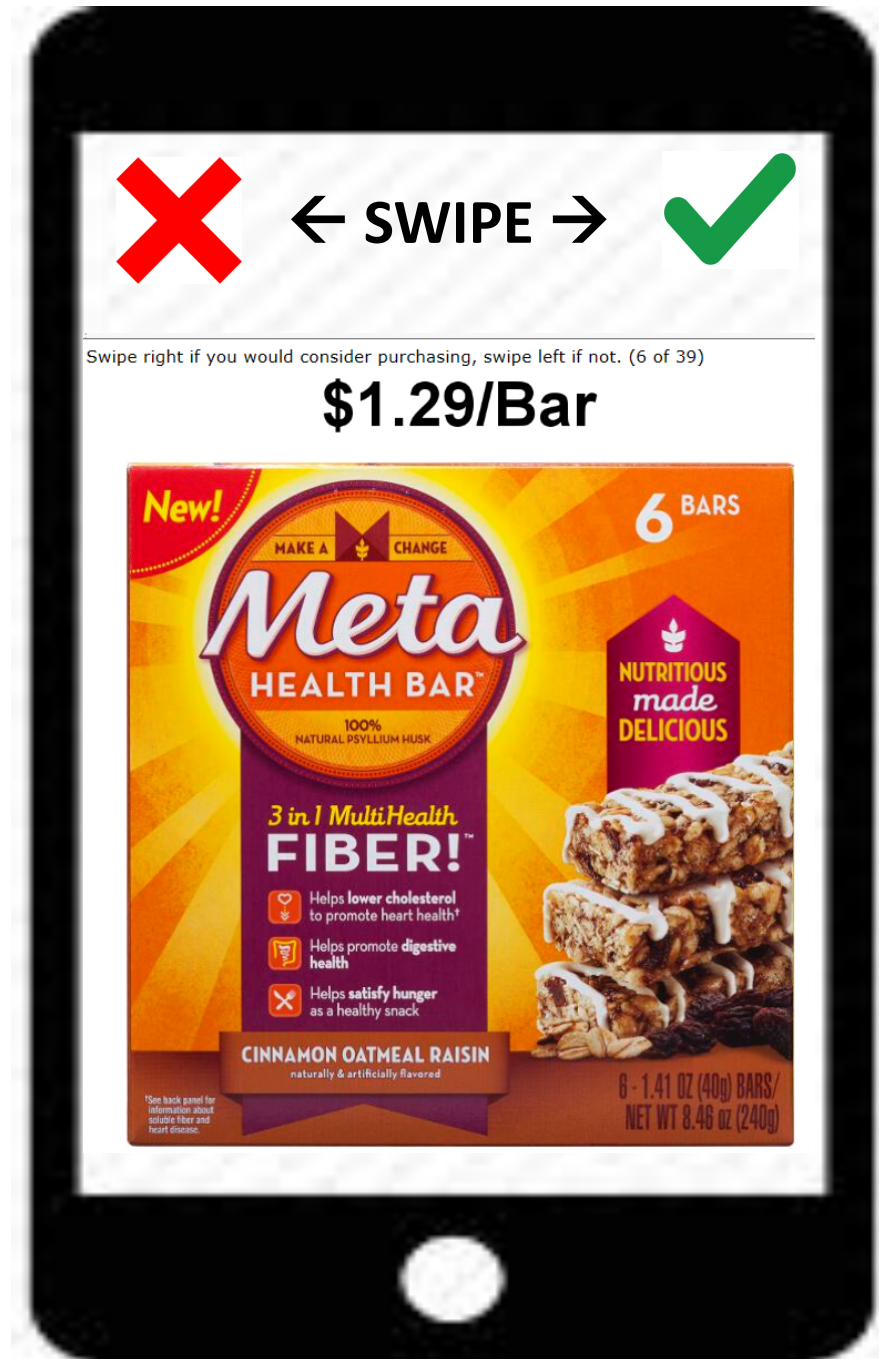
Another detailed example...

**Most**  
**Likely**



# Anchors can weigh you down...

- Pitfall #2 –sequential monadic swipe vs. full list CATA anchors highlight differences in scale
  - Sequential swipe – higher scale (avg. 47% vs. anchor)
  - 40 item CATA list – lower scale (avg. 26% vs. anchor)
- Peak – Standardize the anchor question:
  - Regardless of test design, show 12 items according to “on the fly” rank percentile:
    - 100, 95, 90, 80, 70, ... , 0
    - Shown as “Check all that apply” (comparative)

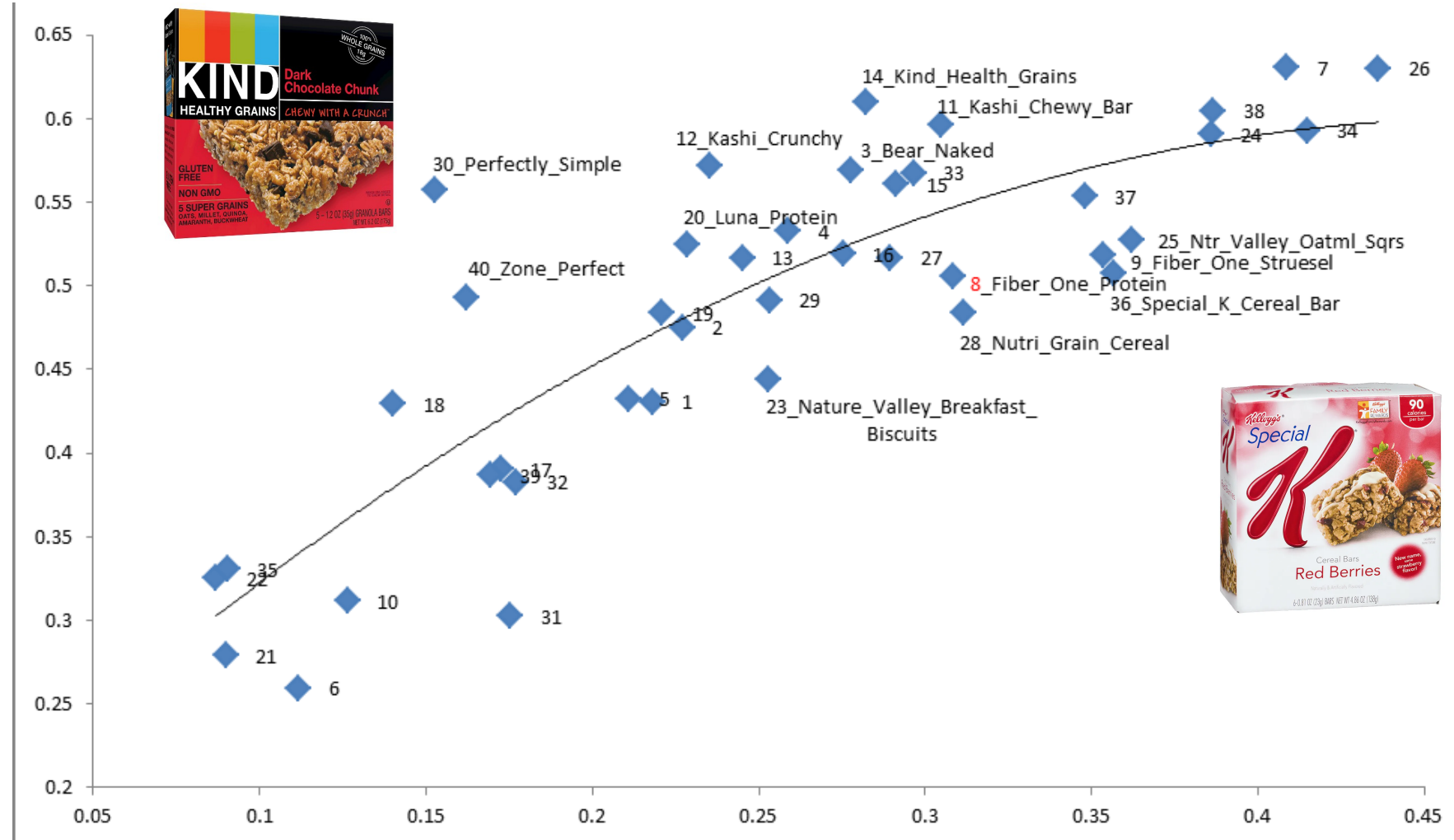




# Get mobile friendly...or else!

- Pitfall #3 – Excluding mobile panelists.
- Misses diversely yuppy parents
- Misses key preference differences
- Baby research study: 90% of young moms entered one survey on a mobile device
- Peak – Develop surveys that work on all devices

Purchase Intent: Mobile vs. PC



# Line extensions dressed for success...

- Pitfall #4: Launching line extensions just to “make news” in the category
  - (Fictitious): “New! Pink Lemonade joins Lemon flavor!”
- Peak: Use MaxDiff to research the unmet consumer need in the category
  - Design set of current competitors, current products, and new (potential) products
  - Package shots + price capture realistic choice share at shelf
  - TURF analysis helps find new products that will maximize incremental share
    - Or, a new line-up to maximize total share

Preview each product by clicking on its thumbnail, then select which product idea you like the most and which you like the least. You'll need to preview all 4 products before answering. (14 of 16)

## Like the Most



## Item 4

**NEW!**

MAKE A CHANGE

# align.

PROBIOTIC SUPPLEMENT

**CHEWABLES**

**24/7 Digestive Support\***  
Clinically Studied *B. infantis* 35624  
naturally helps\*

- Fortify your digestive system with a unique healthy bacteria 24/7\*
- Promote & Support a healthy digestive system
- Maintain digestive balance

Banana Strawberry Smoothie  
Artificially Flavored

**24 CHEWABLE TABLETS**

\*THESE STATEMENTS HAVE NOT BEEN EVALUATED BY THE FOOD AND DRUG ADMINISTRATION. THIS PRODUCT IS NOT INTENDED TO DIAGNOSE, TREAT, CURE, OR PREVENT ANY DISEASE.

**Align available starting at \$26.99 for 24ct.  
Dose 1 chewable per day, \$1.12/day.**



**HELPS YOU FALL ASLEEP NATURALLY**

These statements have not been evaluated by the Food and Drug Administration. This product is not intended to diagnose, treat, cure, or prevent any disease.



**HELPS YOU FALL ASLEEP FAST**

Use as directed for occasional sleeplessness. Keep out of reach of children. Read each label.

# Sensory MaxDiff



- Pitfall #5
  - Some stimuli are too novel, sensory oriented, or difficult to understand virtually
- Peak
  - Run MaxDiff at a central site with live stimuli
    - Scent, tactile, 3D visuals, packages, demos...
    - Adaptive MaxDiff (Orme 2006) limits sensory demands on the panelists

Which scent do you like the most and the least?  
(1 of 15)

	SFK	MYA	MBC	BVT
Like the <u>Most</u>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Like the <u>Least</u>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



# “Hey, I was taken out of context...”

- Pitfall #6 – failure to study in context
  - MaxDiff with plain grid text claims
- Peak – Show stimuli in the intended execution, if known (e.g. on pack)



Which product are you most and least likely to purchase from the Metamucil fiber supplement products below?

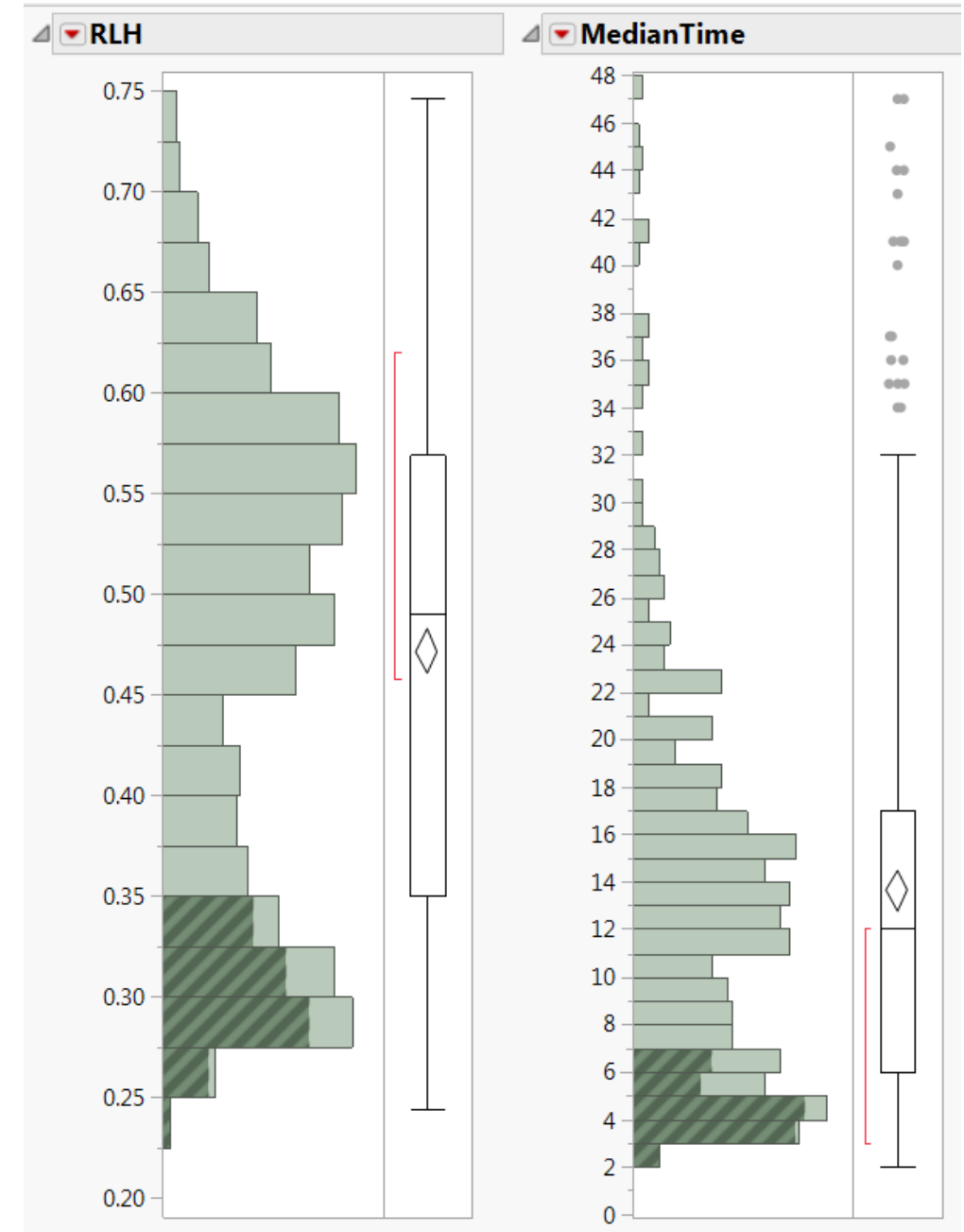
				
	72 DOSES	72 DOSES	72 DOSES	conclusive. 72 DOSES
Most Want	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Least Want	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Different On Pack Claims



# Flat results, flat segments...

- Pitfall #7 – Bad panelists in the mix
  - Flatliners/speeders have flat utilities (low scale), flattening overall results
  - Can produce a huge “flat” segment when clustering raw utility scores
    - Typically, 20% but up to 50% flat segment
- Peaks – Clean & common scaled
  - Filter out speedy, low scale panelists using page times and RLH statistic
    - RLH --> correct prediction probability
    - Hit Rate: from 56.6% to 62.2%
  - “Stretch” the remaining panelists
    - Constrain groups to a common scale in Latent Class analysis, or
    - Cluster (100) interval-scaled utilities



# Too much MaxDiff is never enough...

- Pitfall #8 – “I’ve got 80 claims and I’d like to find the best ones...oh, and my base size is limited...”

1. Vaporize your cold.
  2. The nighttime, sniffing, sneezing, coughing, aching, fever, best sleep with a cold medicine.
  3. Maximum symptom-fighting ingredients to relieve your worst cold symptoms.
- ...
80. Safe cough relief so the family can sleep



- Peak – Adaptive methods show the best items more often
  - Adapt at the population level
    - Bandit MaxDiff (Fairchild, 2015) selects a subset of mostly top overall items to include in the MaxDiff exercise.
    - Boosted Bandit (Orme, 2018) shows the top few overall items 2x more often per respondent.
  - Adapt at the individual level
    - Constructed, Augmented MaxDiff (Bahna, Chapman, 2018) keeps mostly “above the anchor” items
    - Adaptive MaxDiff (Orme, 2006) drops out the “worst” choices per round

"Never say goodbye,  
because saying goodbye  
means going away  
and going away means  
forgetting."  
Peter Pan

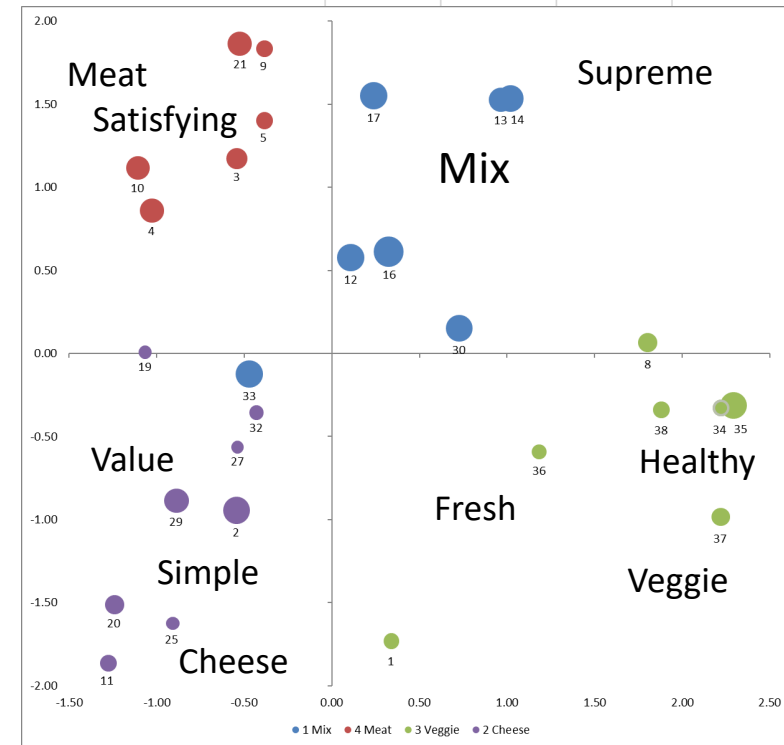


# Love your MaxDiff segments? Never say goodbye again.

- Pitfall #9 – MaxDiff segments are soon forgotten after the project ends
- Peak – MaxDiff produces segments you can really sink your teeth into
  - Design a MaxDiff over a wide ranging list of possible product category benefits
  - Cluster-analyze the choice data
  - PCA plot brings the segments to life
    - Annotate with benefit “themes”
    - Size of bubble = overall score



Ingredient	1 Cheese	2 Mix	3 Meat	4 Veggie
Green peppers	0.13	0.35	0.10	0.46
Mushrooms	0.21	0.48	0.11	0.49
Onions	0.25	0.32	0.09	0.42
Olives	0.13	0.14	0.04	0.34
Red peppers	0.17	0.18	0.02	0.30
Tomatoes	0.42	0.12	0.12	0.37
Pineapple	0.11	0.03	0.25	0.12
Bacon	0.09	0.16	0.39	0.02
BBQ Chicken	0.04	0.12	0.35	0.05
Ham	0.14	0.19	0.41	0.11
Pepperoni	0.16	0.44	0.56	0.07
Sausage	0.08	0.39	0.47	0.04
Plain Cheese	0.58	0.10	0.07	0.27



# Love your MaxDiff segments? Never say goodbye again.

- Search and reapply with the Sawtooth Software typing tool...
  - Find a few MaxDiff questions that best predict cluster membership
  - Ask them in future research to segment or screen new panelists as target consumers

## Set 1 of 3

Most Like	Least Like	
X		Plain Cheese
	X	Italian Sausage
		Onions
		Pepperoni



```
***** Summary of Replications *****
Characteristics of This Run:
  Number of respondents = 662
  Number of tasks = 3
  Items per task = 4
  Replications = 500
  Using all respondents
  Maximizing overall RLH
  Random seed = 555

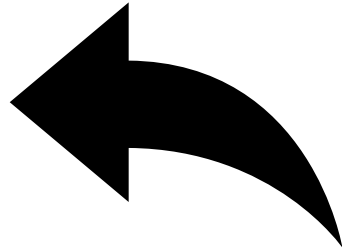
Average Hit Rate Over Replications was 0.8219

The Best Replication was # 192
Best Tasks
  34 13 20 27
  11  9 35  3
  21 34 17 37

Overall Hit Rate =    0.860

Hit Rates for Each Segment
0.898 0.872 0.829 0.842
```






Still need more proof?  
<https://bit.ly/2uwoYwC>



# Step 1 – Ratings Question


How we WANT respondents to behave

 **Best-Worst Scaling (MaxDiff) Example**

Please rate these 12 ice cream flavors.

	1 - Poor	2 - Fair	3 - Good	4 - Very Good	5 - Excellent
Vanilla	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Coconut	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Raspberry	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mango	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cherry	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Neapolitan	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Peach	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Chocolate	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pistachio	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mint Chocolate Chip	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Rocky Road	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Strawberry	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

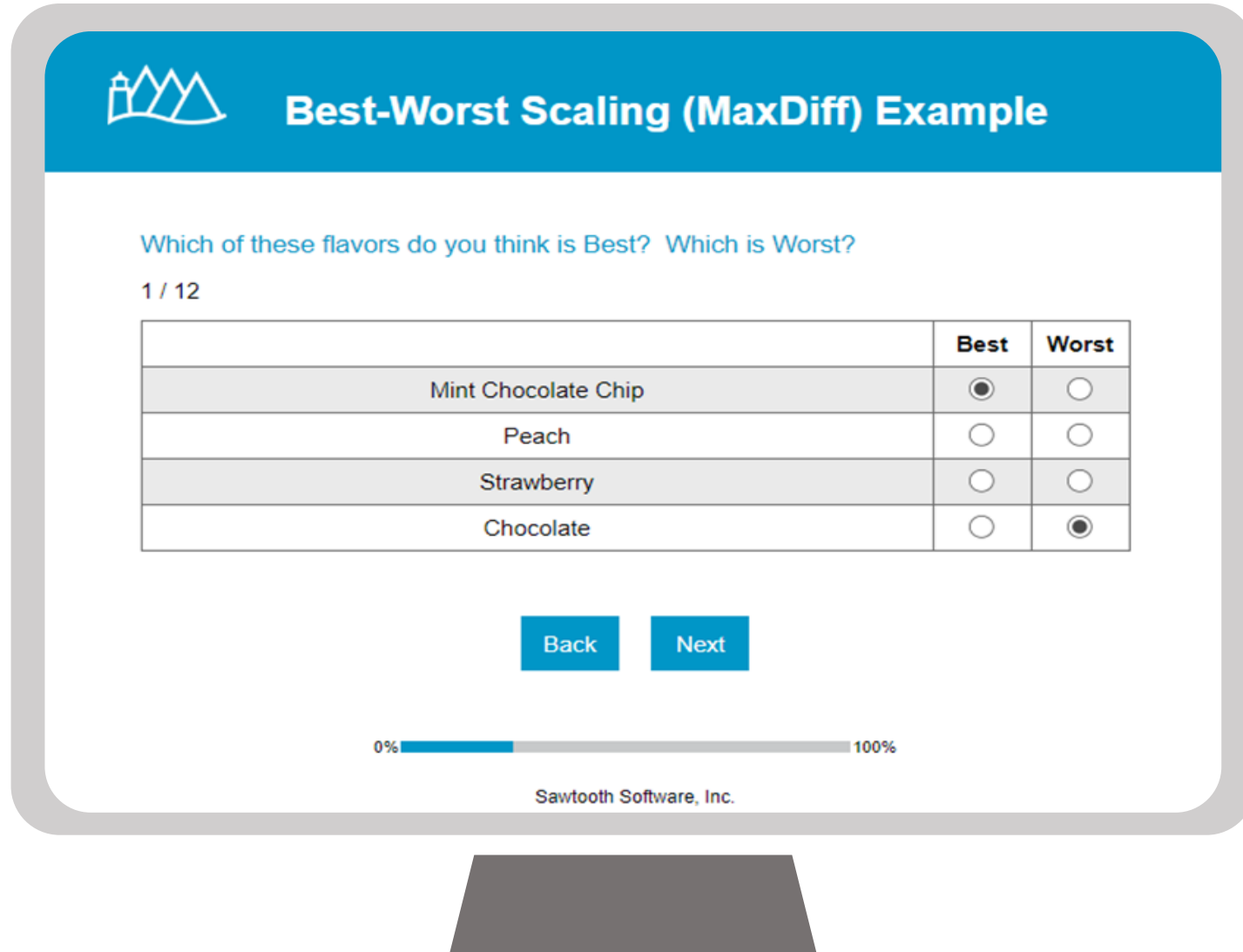
How they ACTUALLY behave...

 **Best-Worst Scaling (MaxDiff) Example**

Please rate these 12 ice cream flavors.

	1 - Poor	2 - Fair	3 - Good	4 - Very Good	5 - Excellent
Mango	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Chocolate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Mint Chocolate Chip	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Vanilla	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Neapolitan	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Raspberry	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Pistachio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Coconut	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Cherry	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Peach	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Rocky Road	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Strawberry	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

# Step 2 - MaxDiff



The desktop view shows a survey titled "Best-Worst Scaling (MaxDiff) Example". It asks the user to select the best and worst flavors from a list. The current question is 1 of 12. The flavors are Mint Chocolate Chip, Peach, Strawberry, and Chocolate. The "Best" column has a radio button selected for Mint Chocolate Chip, and the "Worst" column has a radio button selected for Chocolate. There are "Back" and "Next" buttons, a progress bar at 0%, and the Sawtooth Software, Inc. logo.

**Best-Worst Scaling (MaxDiff) Example**

Which of these flavors do you think is Best? Which is Worst?

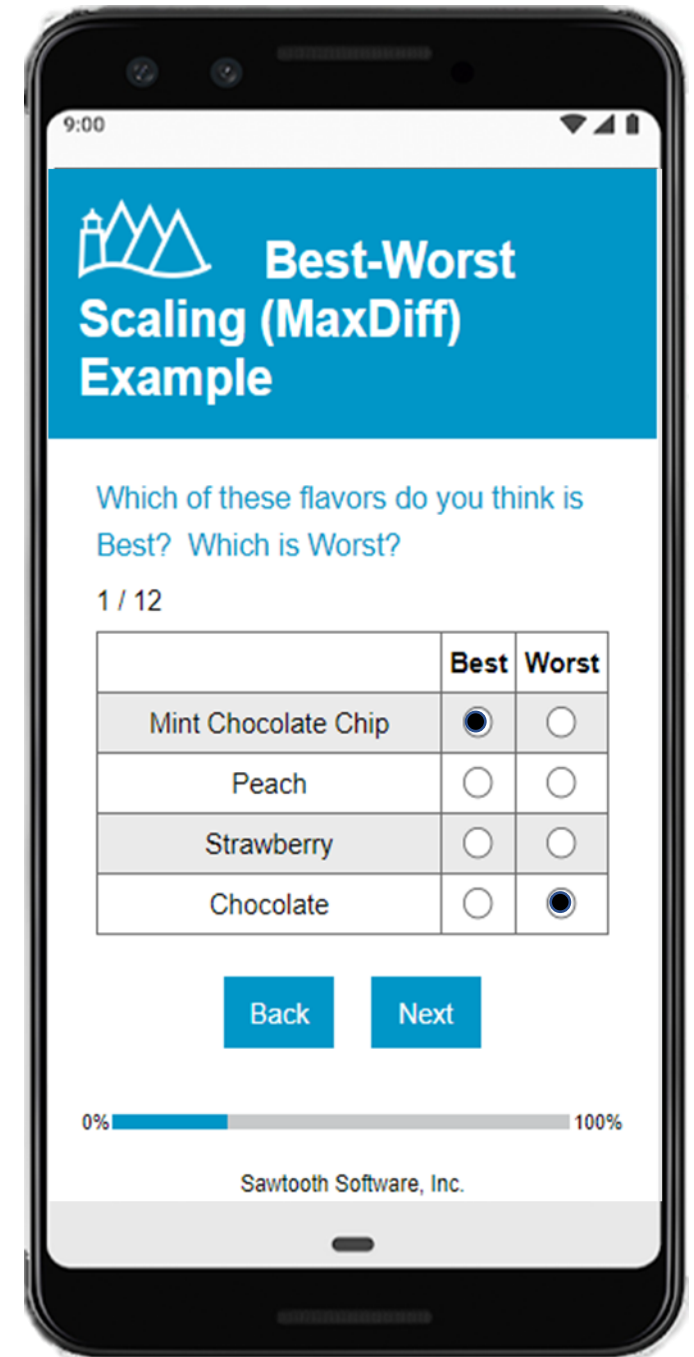
1 / 12

	Best	Worst
Mint Chocolate Chip	<input checked="" type="radio"/>	<input type="radio"/>
Peach	<input type="radio"/>	<input type="radio"/>
Strawberry	<input type="radio"/>	<input type="radio"/>
Chocolate	<input type="radio"/>	<input checked="" type="radio"/>

Back Next

0% 100%

Sawtooth Software, Inc.



The mobile view shows the same survey on a smartphone. The title is "Best-Worst Scaling (MaxDiff) Example". The question is "Which of these flavors do you think is Best? Which is Worst?". It is question 1 of 12. The flavors are Mint Chocolate Chip, Peach, Strawberry, and Chocolate. The "Best" column has a radio button selected for Mint Chocolate Chip, and the "Worst" column has a radio button selected for Chocolate. There are "Back" and "Next" buttons, a progress bar at 0%, and the Sawtooth Software, Inc. logo.

**Best-Worst Scaling (MaxDiff) Example**

Which of these flavors do you think is Best? Which is Worst?

1 / 12

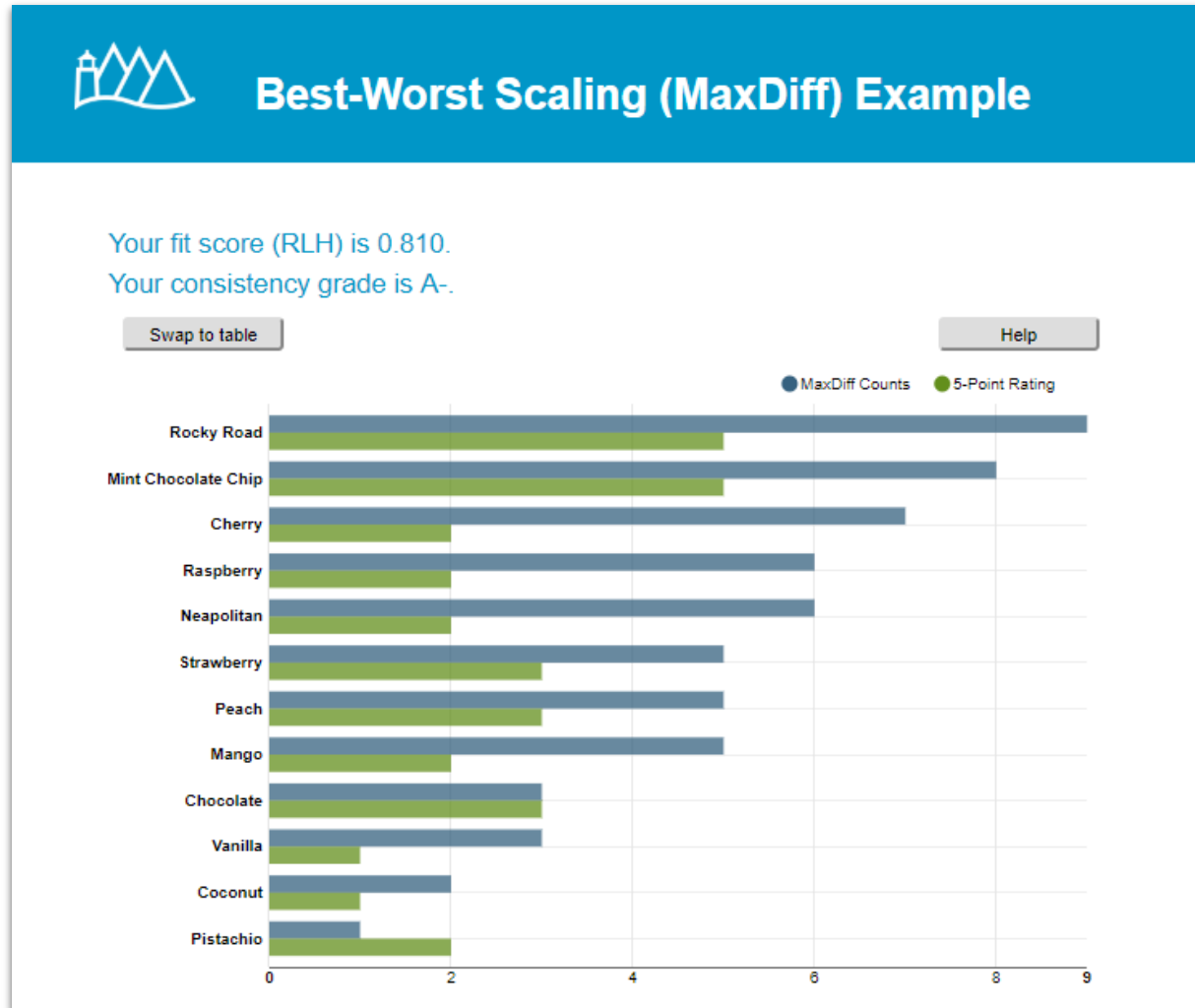
	Best	Worst
Mint Chocolate Chip	<input checked="" type="radio"/>	<input type="radio"/>
Peach	<input type="radio"/>	<input type="radio"/>
Strawberry	<input type="radio"/>	<input type="radio"/>
Chocolate	<input type="radio"/>	<input checked="" type="radio"/>

Back Next

0% 100%

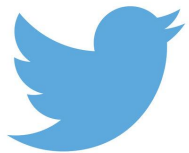
Sawtooth Software, Inc.

# Step 3 – Compare your results





# Questions?



@MeganPeitz  
@SawtoothSoft