BEYOND MAXDIFF

Taking Your MaxDiff Survey to the Next Level

Aaron Hill, Vice President
Sawtooth Software, Inc.
aaron@sawtoothsoftware.com

RESEARCH NEEDS

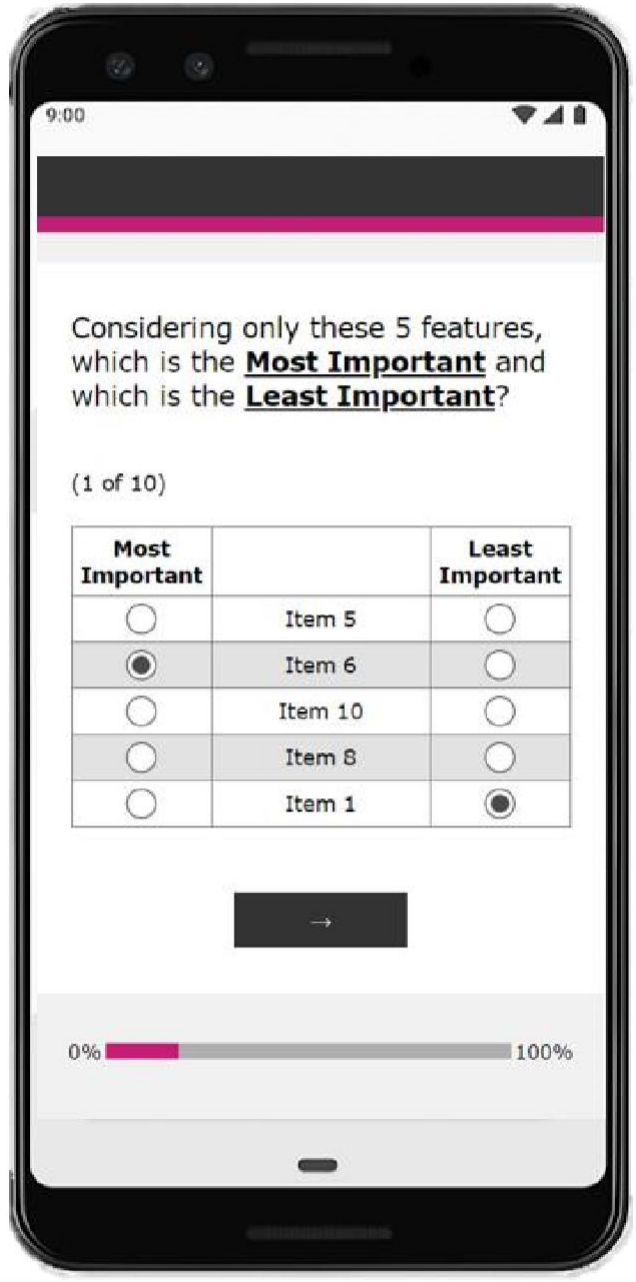
Test brand messaging
Prioritize feature development
Select best packaging
Flavor/scent/color selection
Meaningful line extensions
Political positions
Employee benefits optimization

ONE TOOL IN OUR ARSENAL?
MAXDIFF

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

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, brands, features, packaging options, vacation destinations, and more!

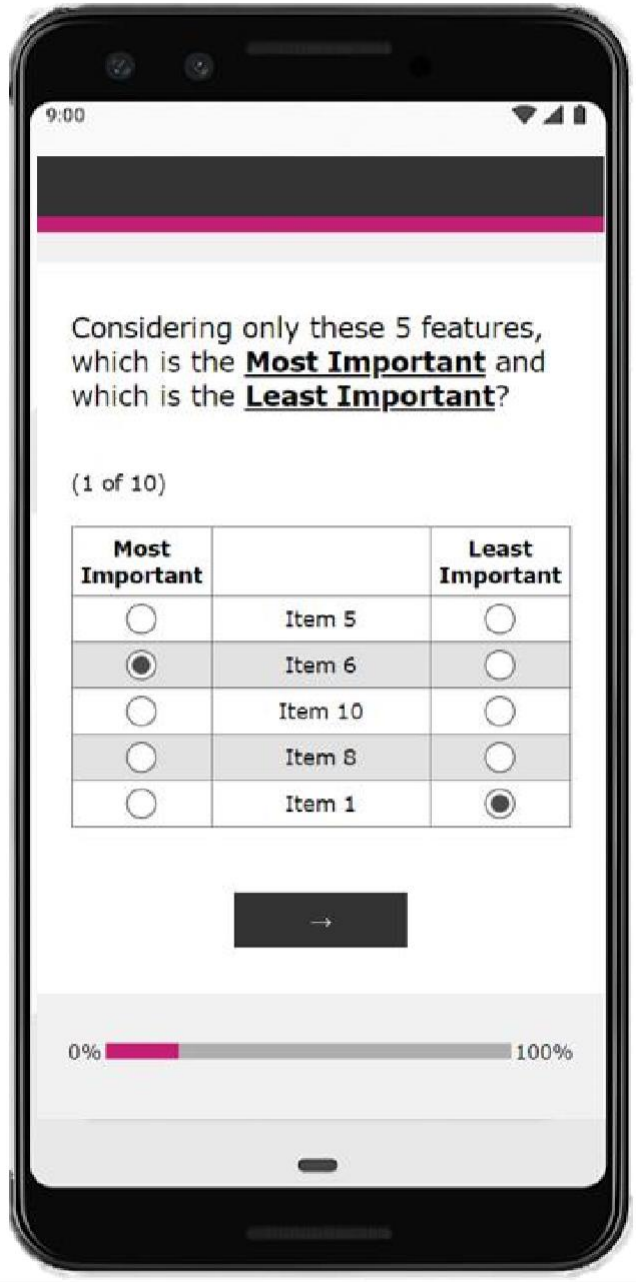


Quick Review

- Respondents typically shown 2-6 items at a time, asked to indicate extreme items (best/worst, most important/least important, etc.)

- Task is repeated many times, showing a different set of items in each task

- Usually 20 or so items in total



Please consider how important different features are when selecting a movie to attend.

Considering only these five features, which of these features makes you Most Likely to Attend a movie, and which makes you Least Likely to Attend?

	Most Likely to Attend	Least Likely to Attend
The film features a major star	○	○
The film won an Academy Award	○	○
The film features lots of action sequences, including car crashes	○	○
The film is based on a bestselling historical fiction	○	○
The film is about a girl and her horse	○	○

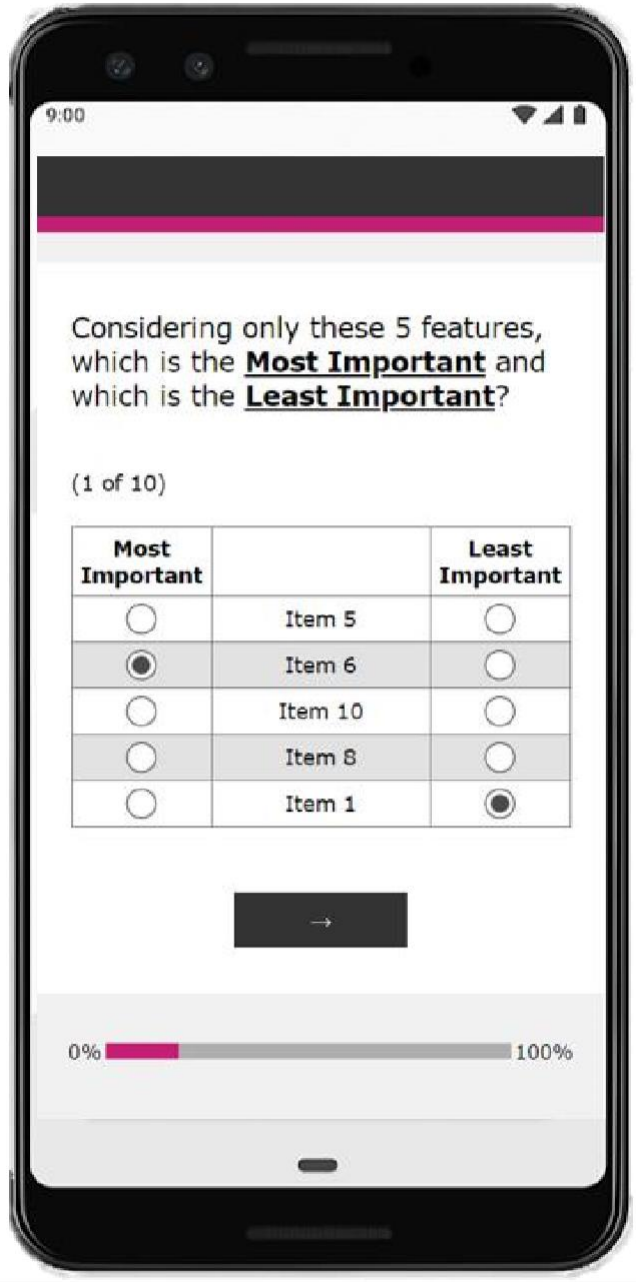
Click the 'Next' button to continue...

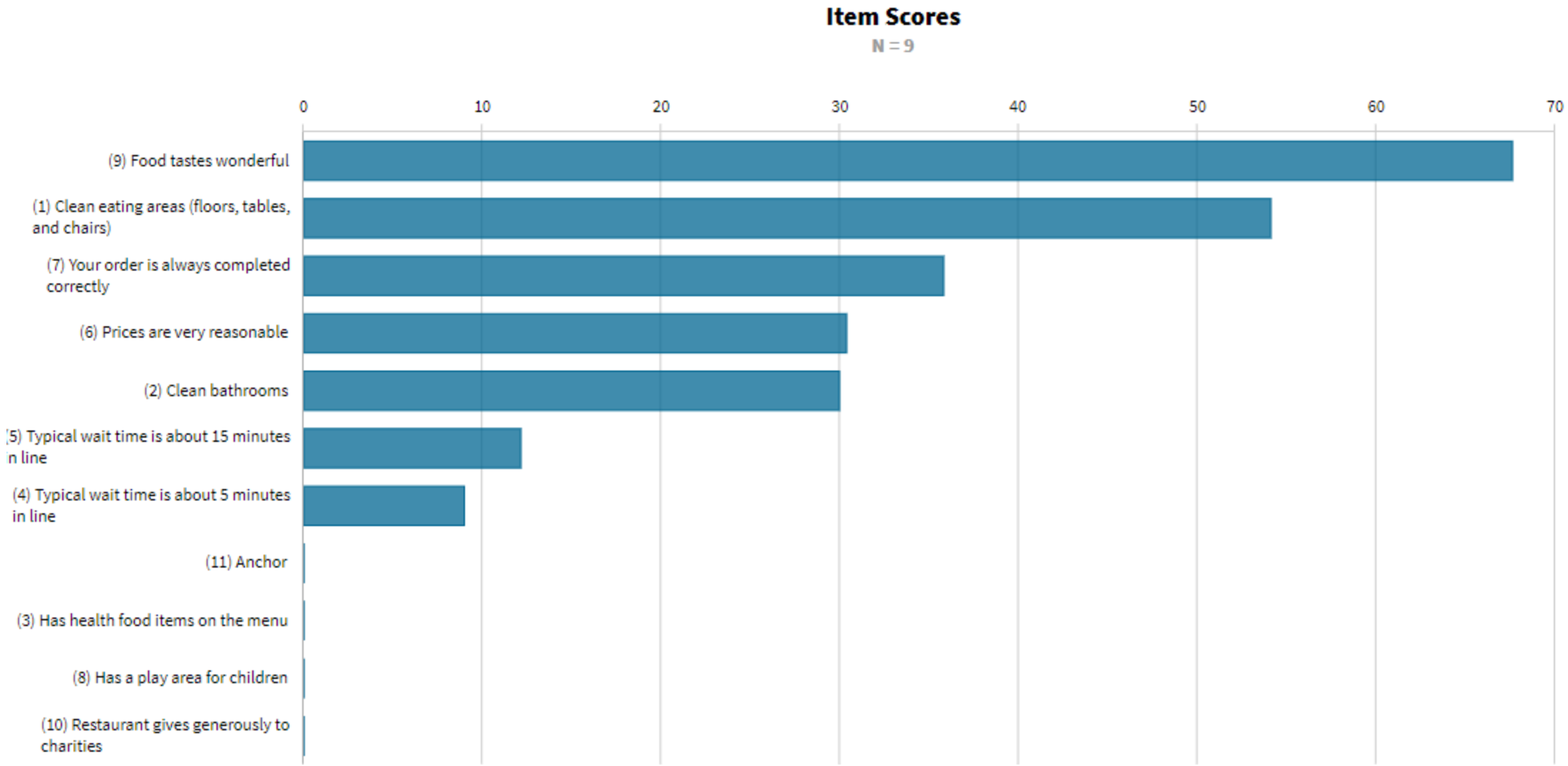
Next

Quick Review - Analysis

- Data are scored using counts analysis or by applying a logistic regression to find each item's relative weight.

- Resulting model provides rank order data, but also more precise ratio-scaled scores for each item.




Item Scores
N = 9
0
10
20
30
40
50
60
70
(9) Food tastes wonderful
(1) Clean eating areas (floors, tables, and chairs)
(7) Your order is always completed correctly
(6) Prices are very reasonable
(2) Clean bathrooms
5) Typical wait time is about 15 minutes n line
(4) Typical wait time is about 5 minutes in line
(11) Anchor
(3) Has health food items on the menu
(8) Has a play area for children
(10) Restaurant gives generously to charities

WHY DO WE LIKE IT?

MaxDiff Data > Ratings Data

Greater discrimination among items

Greater discrimination between respondents on the items

No scale bias

Fits easily into mobile interviews

MaxDiff Data > Rankings Data

Ranking is impractical with 10+ items

Ordinal scaled results

What do you report? Top Rank? Average Rank? Top 2?

WHEN DOES IT WORK WELL

- Need to score typically 6 - 20 somewhat unrelated items
  - If you're interested in ***combinations of items***, use conjoint analysis instead
- Requires decent sample size
- List typically must be exhaustive and limited in scope
  - Not a good tool for "blue sky" ideation, except to evaluate proposed ideas

STANDARD MAXDIFF IS GREAT, BUT IT DOESN'T ALWAYS MEET OUR NEEDS

- What about when we have ***detailed visuals*** instead of text?
- Or even worse, what do I do if I have items that need to be ***touched, smelled, or tasted***?
- How do I know if an idea is ***actually good***, not just better than other bad ideas?
- What if my question is, "***Which set of products*** should I carry," not, "Which product is best"?
- What if I have ***large lists*** of items?
- Can I use it as part of a ***broader research project***?

PROBLEM 1: WE NEED IMAGES, NOT TEXT

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

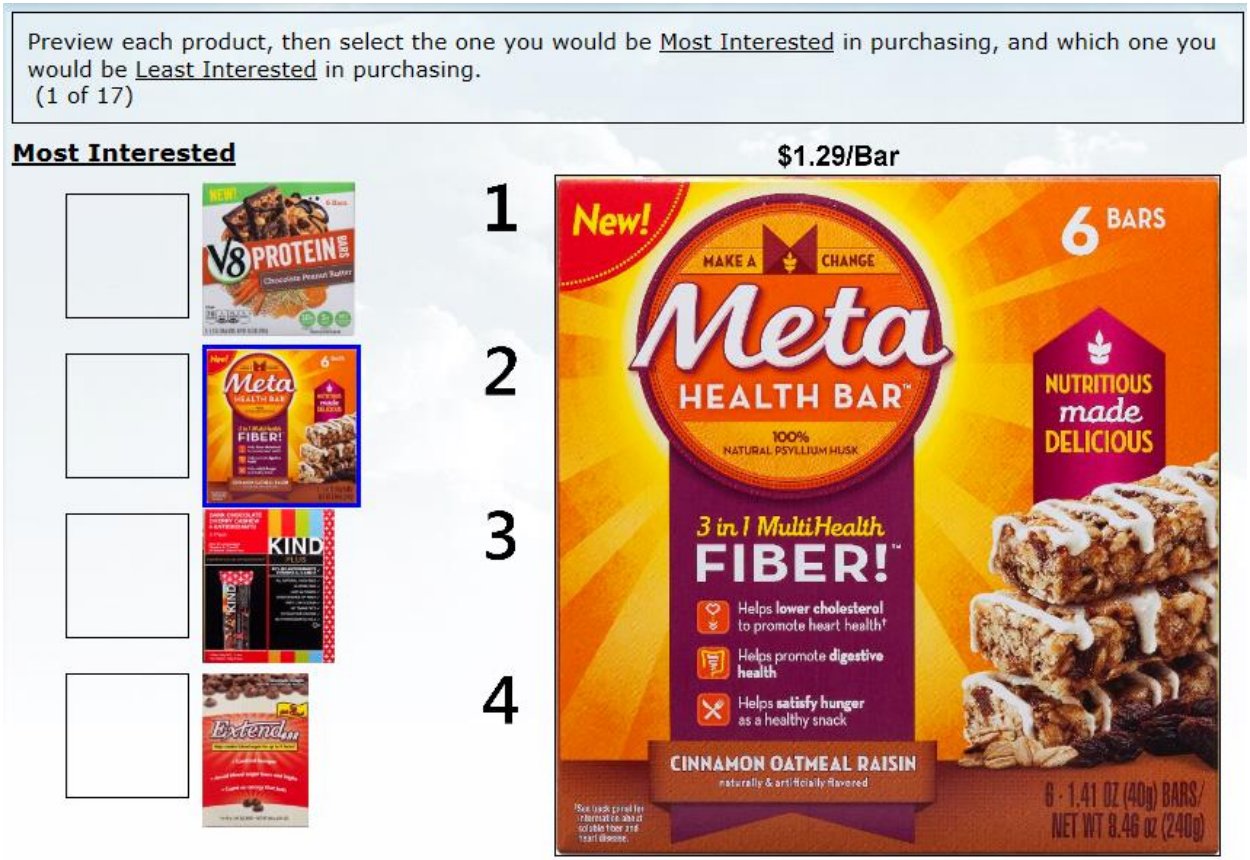
Option 1	Option 2	Option 3	Option 4
			
Mad Angles	Tedhe Medhe	Tangles	Yumitoes
100 gm	125 gm	150 gm	200 gm
Rs.7	Rs.10	Rs.15	Rs.20
☐	☐	☐	☐

Most Prefer		Least Prefer
○		○
○		○
○		○
○		○

PROBLEM 1: IMAGES INSTEAD OF TEXT

Solution: Panelist previews all thumbnail images which launch (near) full-screen images

- Then panelists can review them and make best/worst choices.



IMPORTANCE OF CONTEXT

Sometimes, we do the right research but don’t provide the necessary context.

Which of these claims appeals to you most (best), and which appeals least (worst)?

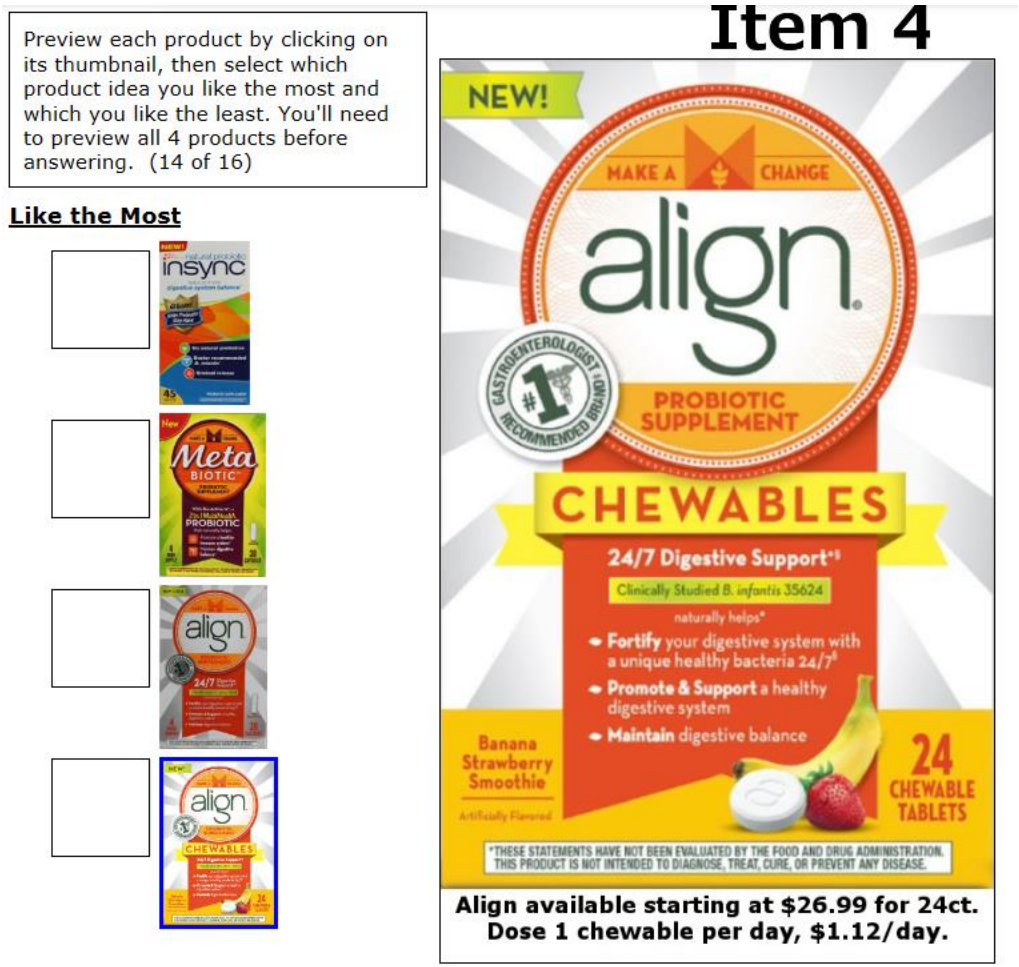
Best	Worst	Claim
○	○	Helps maintain healthy blood sugar levels
○	○	Helps you feel less hungry between meals
○	○	Helps lower cholesterol to promote heart health
○	○	Promotes digestive health

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

Most *Want*				
Least *Want*				

IMPORTANCE OF CONTEXT

- Use MaxDiff to research unmet consumer needs in the category
  - Design set of current competitors, current products, and potential new products
  - Package shots + price capture realistic choice share at shelf



APPLYING TURF ANALYSIS TO OPTIMIZE PRODUCT LINEUP

TURF: Total Unduplicated Reach & Frequency

- TURF analysis helps find new products that will maximize incremental share, or a new lineup to maximize total share.

Items in Portfolio	Reach	Frequency (n=450)
1 & 13 & 7 & 22	76%	484
1 & 13 & 9 & 4	73%	469
13 & 7 & 22 & 9	68%	443
1 & 9 & 23 & 13	67%	429
1 & 9 & 22 & 4	63%	428

PROBLEM 2: SENSORY WORK

Some stimuli are too novel, sensory oriented, or difficult to understand virtually

Which scent do you like the most and the least?

(1 of 15)

	SFK	MYA	MBC	BVT
Like the Most	○	○	○	○
Like the Least	○	○	○	○

PROBLEM 2: SENSORY WORK

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

PROBLEM 3: “ARE ALL MY IDEAS GOOD IDEAS?” – ANCHORING MAXDIFF

MaxDiff scores tell you which items are preferred, but not whether they are actually acceptable or not.

Solution: Anchor your MaxDiff by including additional questions

- Choose all that apply: “Which of these flavors would you buy?”
- Mobile-friendlier options like “Swipe left/right”

NO ANCHOR

Item 1

Item 3

Item 10

Item 9

Item 2

Item 7

Item 6

Item 5

Item 4

Item 8

We now know the preference order, but which of these are good ideas and which are bad ideas?

NO ANCHOR

Item 1
Item 3
Item 10
Item 9
Item 2
Item 7
Item 6
Item 5
Item 4
Item 8

We now know the preference order, but which of these are good ideas and which are bad ideas?

We ask!

Which of these flavors would you consider purchasing? (Select all that apply)

- ❑ Item 1
- ❑ Item 3
- ❑ Item 10
- ❑ Item 9
- ❑ Item 2
- ❑ Item 7
- ❑ Item 6
- ❑ Item 5
- ❑ Item 4
- ❑ Item 8

NO ANCHOR

Item 1
Item 3
Item 10
Item 9
Item 2
Item 7
Item 6
Item 5
Item 4
Item 8

We know the preference order, but which of these are acceptable and which are bad ideas?

WITH ANCHOR

Item 1
Item 3
Item 10
Item 9

Good ideas

Item 2
Item 7
Item 6
Item 5
Item 4
Item 8

Not-so-good ideas

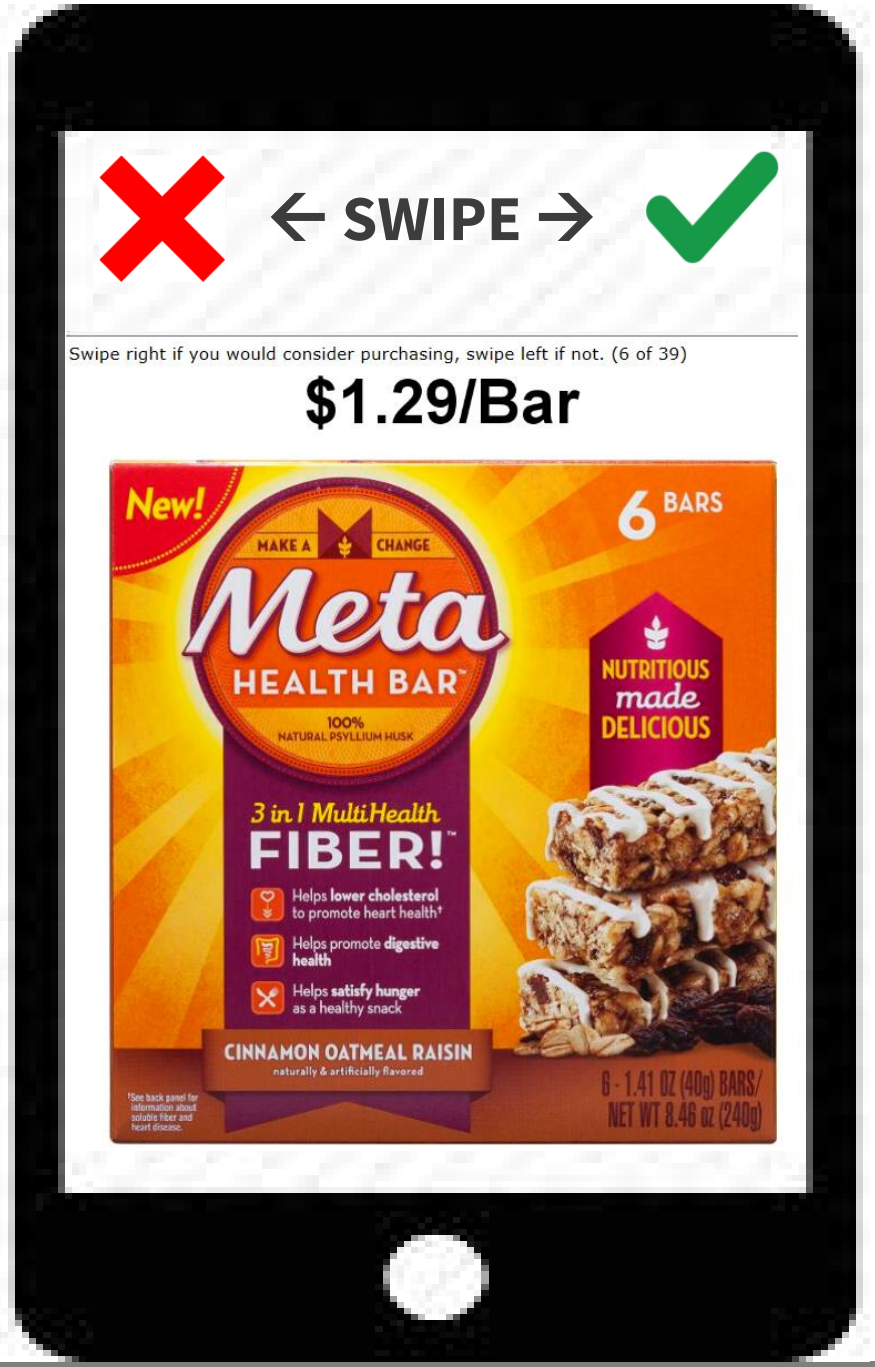
SOLVED ONE PROBLEM, ADDED ANOTHER…

Now we have differences in scale

- Sequential swipe – higher scale (avg. 47% vs. anchor)
- 40 item CATA list – lower scale (avg. 26% vs. anchor)

Solution: 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)



NEXT PROBLEM: WHAT DO I DO IF I HAVE 80+ ITEMS, NOT JUST 20?

1. Vaporize your cold.
2. The nighttime, sniffling, 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

SOLUTIONS

Sparse MaxDiff – show each item less often (maybe even lower than 1x each) and rely on large respondent numbers to compensate

Express MaxDiff – show each respondent a small random subset of the items (say, 20 out of 100) and then “borrow” from other respondents to fill in the gaps

Bandit MaxDiff – like Express MaxDiff, but instead of randomly drawing items, we use past respondent data to pick from winning concepts

Adaptive MaxDiff – Start with all the items, then only carry the winners forward

SOLUTION: ADAPT YOUR MAXDIFF

Adaptive methods show the preferred 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.

SOLUTION: ADAPT YOUR MAXDIFF

Adaptive methods show the preferred items more often

- 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

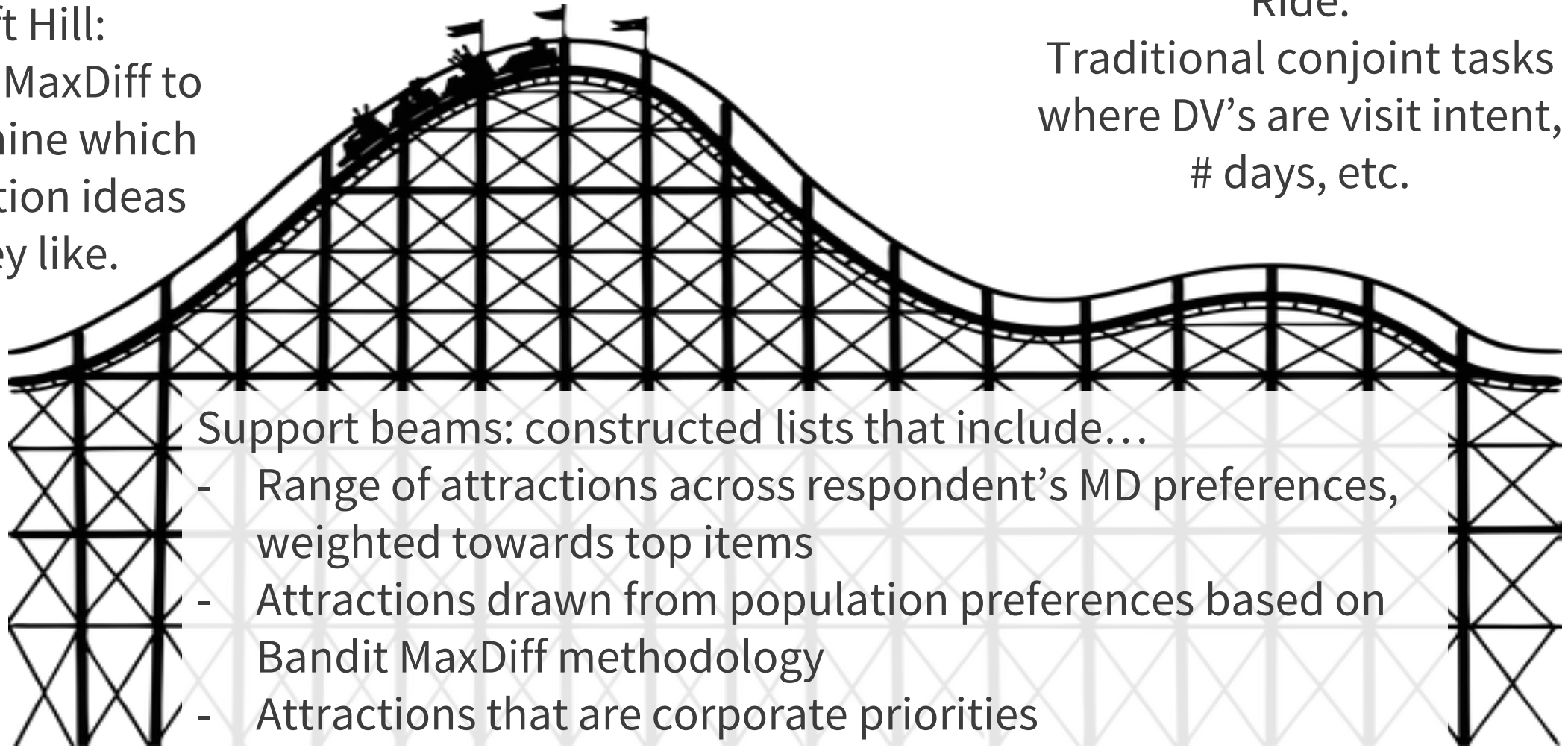
FINALLY

Can I use MaxDiff in the context of a bigger experiment?

THEME PARK RIDE OPTIMIZATION

Lift Hill:
Sparse MaxDiff to determine which attraction ideas they like.

Ride:
Traditional conjoint tasks where DV's are visit intent, # days, etc.



Support beams: constructed lists that include…

- Range of attractions across respondent's MD preferences, weighted towards top items
- Attractions drawn from population preferences based on Bandit MaxDiff methodology
- Attractions that are corporate priorities

CONCLUSIONS

MaxDiff can be a powerful tool in understanding consumer preferences.

Methodology can be adapted to fit unique situations: bigger problems, non-textual items, etc.

Results can be used to provide reproducible, meaningful segments, differentiated ratio-scaled scores, and portfolio optimization.

QUESTIONS?

Contact Aaron Hill
aaron@sawtoothsoftware.com
801-477-4700