



# Should We Be Re-Thinking Product Launch Success Hurdles?

Sensometrics Virtual Online Conference  
October 2020



Patti Wojnicz, P&K Research



Juan Martinez, P&K Research



**P&K**  
research



# 85% of New Product Launches Fail

As statisticians and product researchers, we know this is a concerning percentage. Maybe its time to rethink product performance action hurdles and how we define launch readiness



# What are the right action hurdles?

## Typical Action Standards are based on

- A small set of key after use performance indicators (PI, overall liking, expectations)
- Aggregate data that average over individuals
- One context only (e.g. branded CLT)

## Here we address the following questions

- What can we learn from considering the results of more than one context, and the consistency in ratings by individuals across contexts
- Instead of just post-use KPI's, what can we learn from comparing pre-use to post-use scores  
e.g. the difference between concept appeal and post use product appeal

# What started us thinking about consistency across context ...

- In 2017, P&K Research partnered with Ohio State University
- The same consumers (N=63) evaluated 4 red wines in 3 different contexts/environments



Traditional Controlled



Virtual Immersive



Actual Wine Bar

Data in aggregate (total), revealed the same overall liking rank order across all 3 contexts; no significant context effect



I like both, but the context drives which one I like more



However, at the individual level, context did have an effect. The rank order of wines changed across contexts, but not consistently for all consumers.

While all wines were liked the same in aggregate, we determined that the conclusion needs qualification.

Maybe some wines are better suited for some contexts than others?

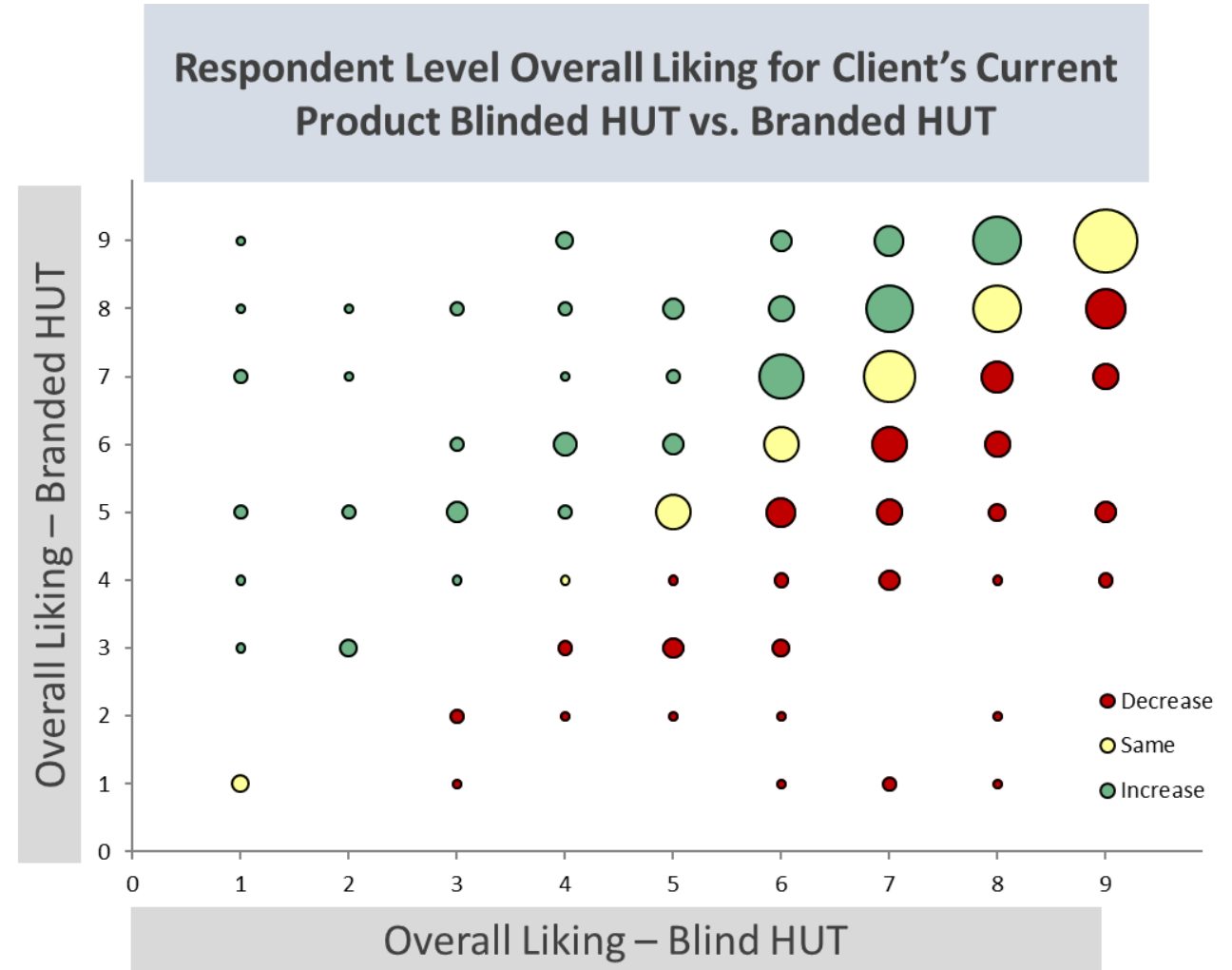
\*2018 Hannum et Al

**Product  
Performance Across  
Blinded vs.  
Branded  
Conditions**

We analyzed N = 34 products across 5 studies (different categories), where the same respondent evaluated the same product across 2 conditions

*(Blinded, Day 1)*  
*(Branded, Day 2)*

Example



# We considered two types of overall liking rating consistency across conditions: Exact Same and Same Tier (T3B / M3B / B3B)

## Cross Tab of Individual Liking (9-Point Scale) Across Both Conditions...

### Exact Same

		Overall Liking - Branded									
		1	2	3	4	5	6	7	8	9	
Overall Liking - Unbranded	1	Dislike extremely	0	0	1	0	0	1	0	0	0
	2	Dislike very much	0	0	0	0	1	0	0	0	0
	3	Dislike moderately	0	0	0	0	0	0	1	1	0
	4	Dislike slightly	0	0	1	0	1	0	3	2	0
	5	Neither like nor dislike	0	0	0	0	1	0	2	2	0
	6	Like slightly	1	0	1	1	1	8	1	6	2
	7	Like moderately	1	0	2	1	2	3	7	11	1
	8	Like very much	0	1	0	1	3	0	4	7	4
	9	Like extremely	0	0	0	0	1	0	2	2	5

### Within the Same Tier (T3B / M3B / B3B)

		Overall Liking - Branded									
		1	2	3	4	5	6	7	8	9	
Overall Liking - Unbranded	1	Dislike extremely	2	0	0	1	0	1	0	0	0
	2	Dislike very much	0	0	0	1	0	0	0	0	0
	3	Dislike moderately	0	0	1	2	1	2	2	0	0
	4	Dislike slightly	0	0	0	2	2	1	1	4	2
	5	Neither like nor dislike	0	0	0	0	0	2	1	3	1
	6	Like slightly	0	1	1	3	3	10	6	5	2
	7	Like moderately	0	0	1	3	2	8	10	7	3
	8	Like very much	0	0	2	1	0	4	7	13	7
	9	Like extremely	0	0	0	0	1	2	3	4	12

OL Decreased for Individual

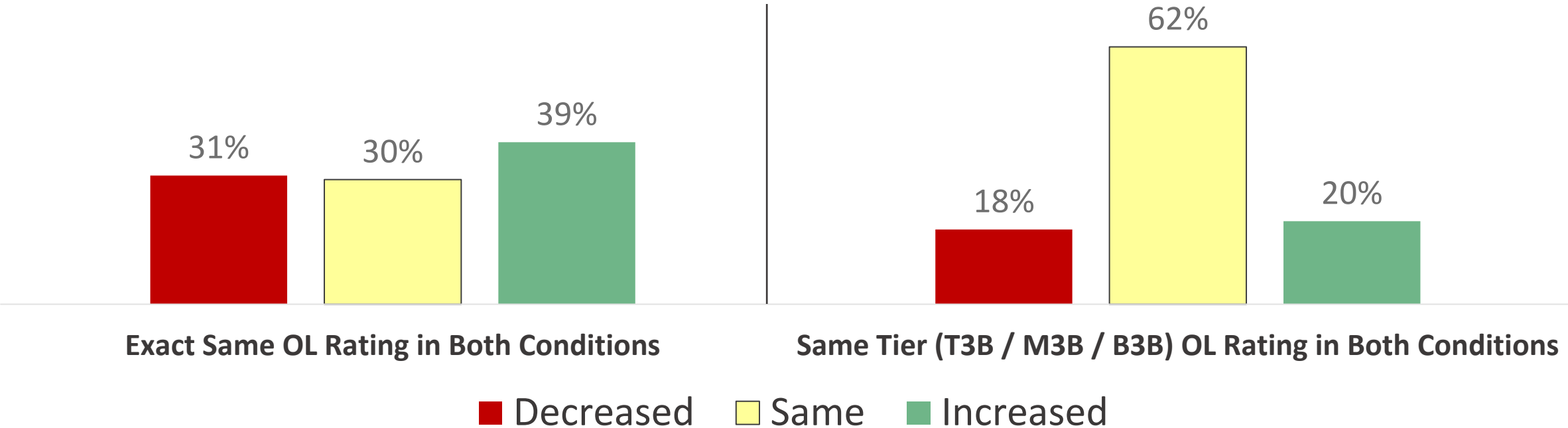
Individual OL Stayed Same  
(Exact or Same T3B / M3B / B3B Tier)

OL Increased for Individual



# Individual overall liking ratings remain the exact same ~1/3 of the time, whereas ~2/3 of the time they are within the same tier (T3B / M3B / B3B)

Individual Shift in Overall Liking Rating Across Conditions (Blind vs. Branded)\*

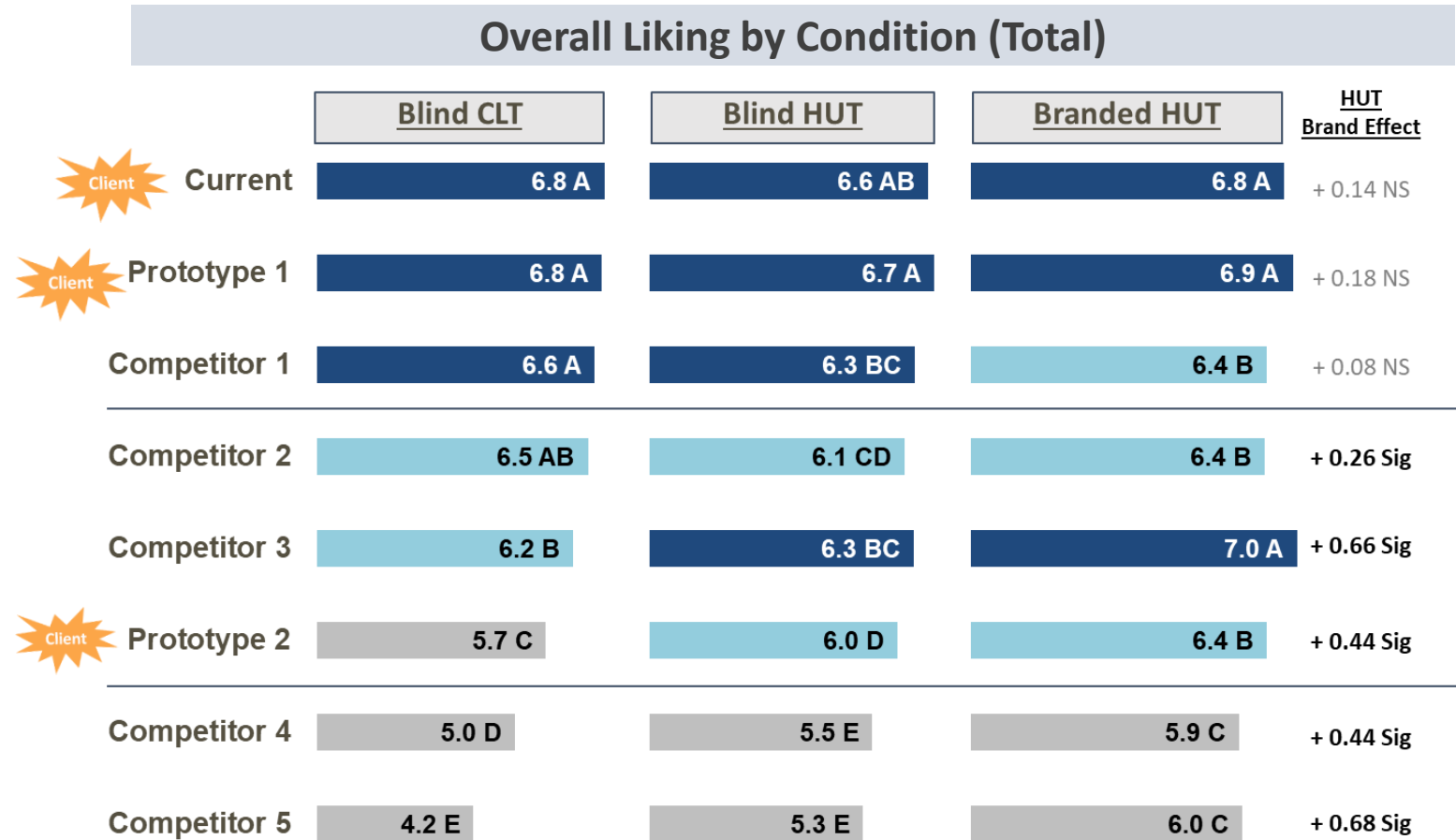


\*Average across N=34 products from 5 studies



# If we were to consider product performance among consistent raters (however defined), would we make different decisions? Applying this theory to one case where the same products were evaluated across 3 conditions...

- Similar to results from the wine study, in total, condition does not influence rank order of performance based on overall liking
- The best and worst liked products remain consistent across all 3 conditions

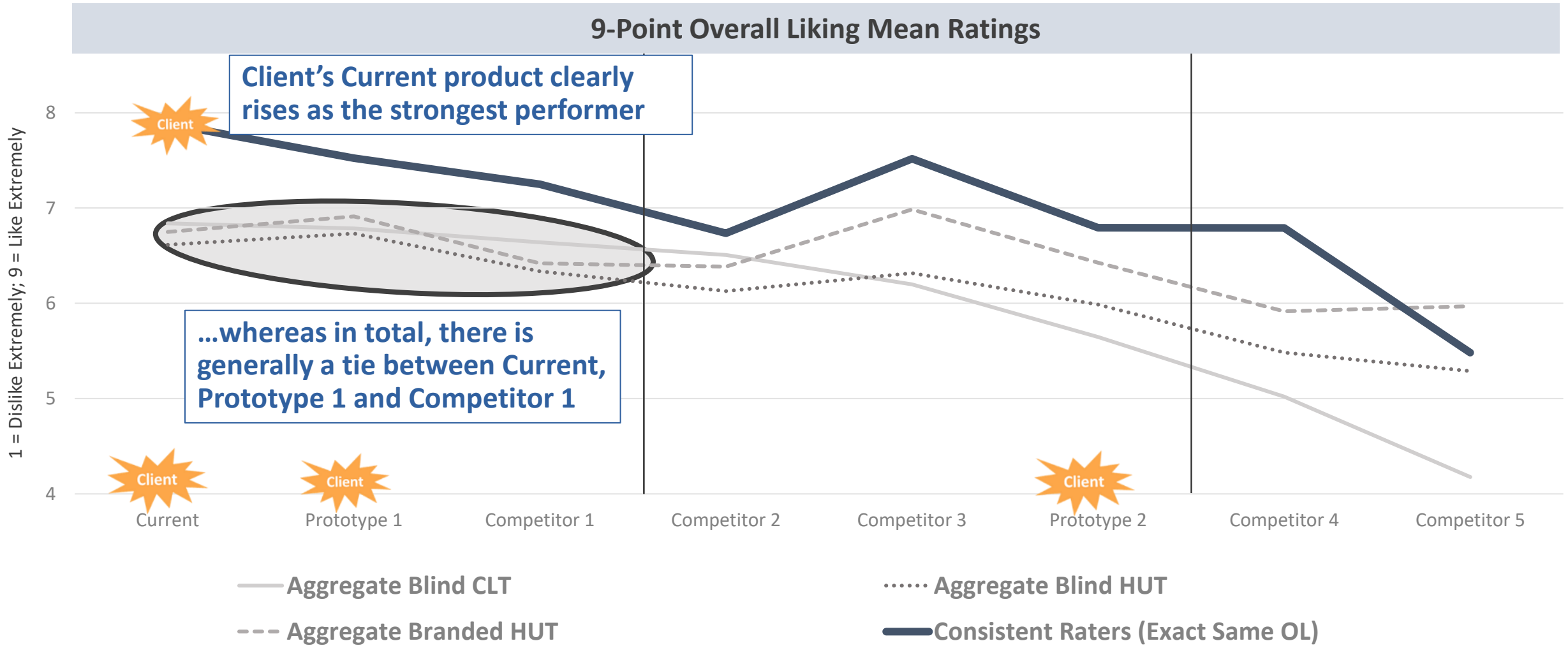


*Products ranked highest to lowest based on CLT Overall Liking*  
 Annotation using common lettering: Capital Letters indicate Significant Difference @ 95% C.L.

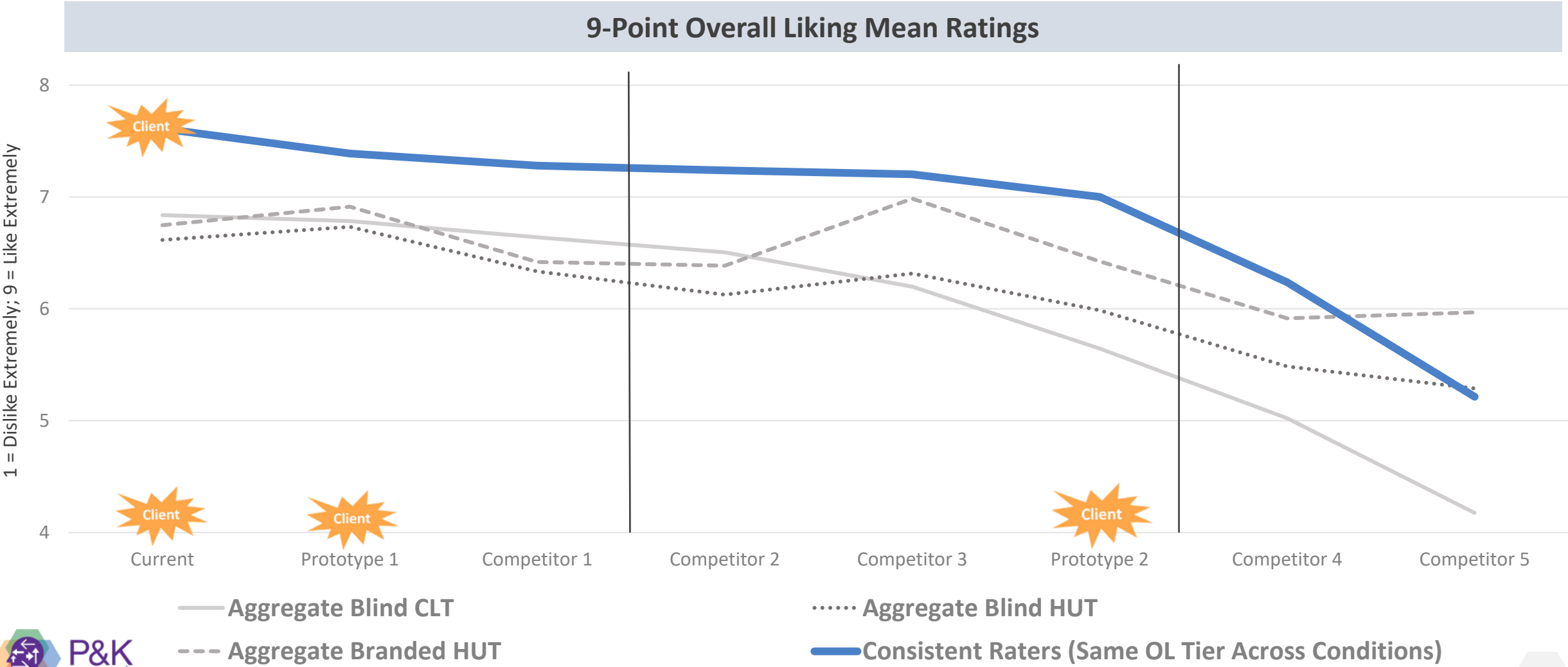
N = 300

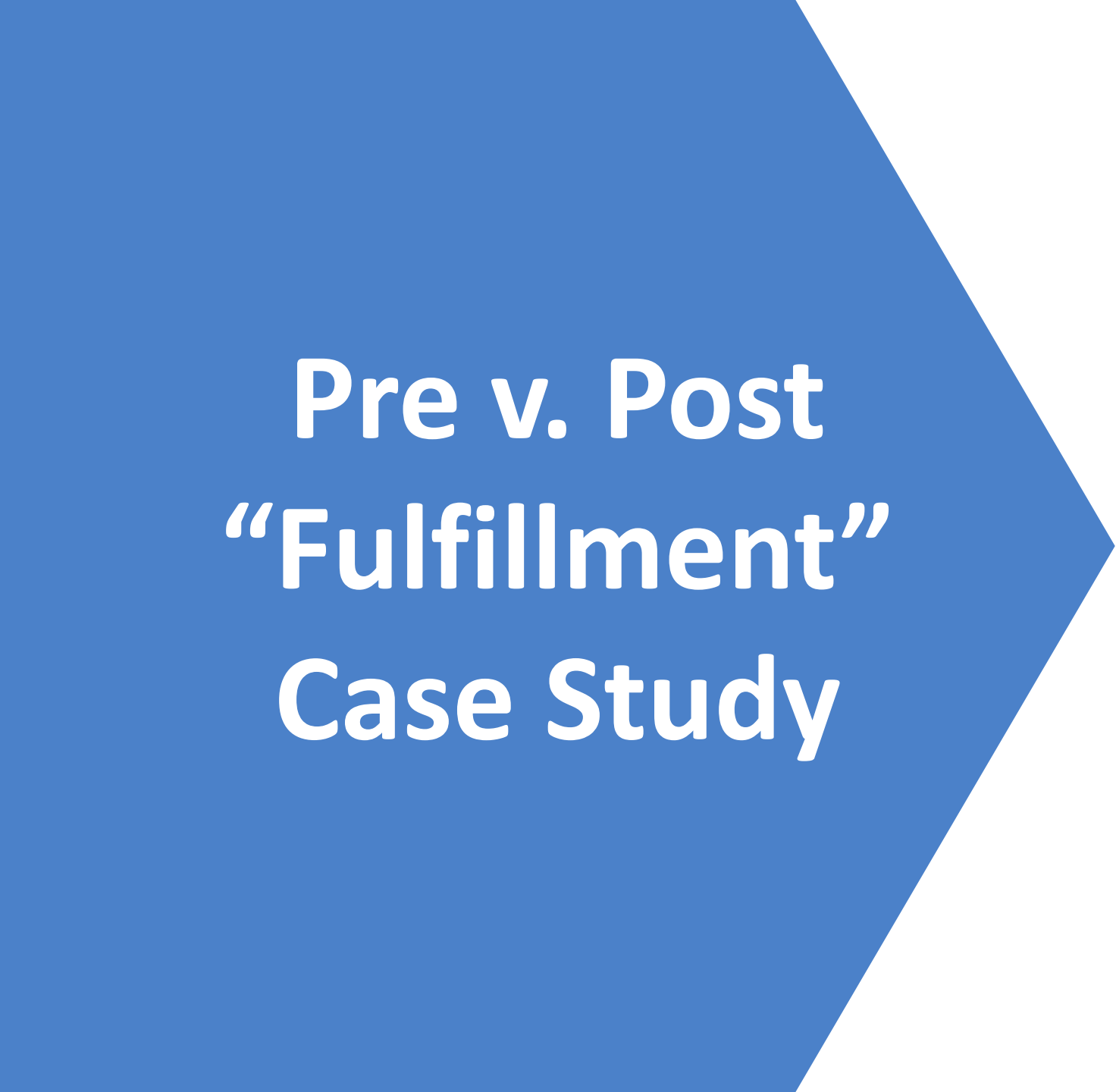


# Analyzing data among those who gave the Exact Same OL for the product across all three conditions...



# Among the Same OL Tier raters, absolute ratings are higher, and Current remains the best liked. Analysis could focus on what is driving an increase or decrease in ratings at the individual levels.



A large blue arrow pointing to the right, which serves as a background for the title text.

# Pre v. Post “Fulfillment” Case Study

# Case Study: Concept vs. Product Fulfillment

- Typical screening criteria was employed:

**100%**  
**Concept**  
**Acceptors**  
**T3B Purchase**  
**Intent**  
Definitely / Probably / Might Buy



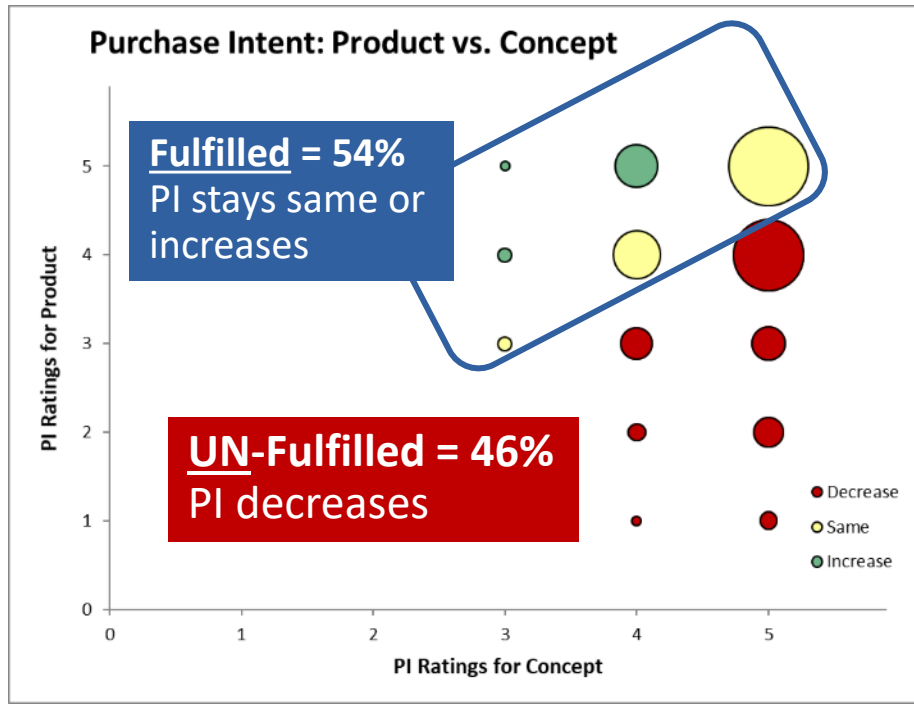
# Case Study: Concept vs. Product Fulfillment

- According to typical action standard hurdles – *In Aggregate* – Team met the first of their two hurdles: **≥ 75% T2B Post Use Purchase Intent**



# Case Study: Concept vs. Product Fulfillment

- However, analyzing data at the individual level revealed lack of consistency across conditions
- **Hurdle = 67% Fulfillment / only 33% NOT fulfilled** by comparing concept vs. product purchase intent for each consumer individually



**Did NOT Meet Action Standard**



# Case Study: Concept vs. Product Fulfillment

- **Analysis\*** then focused on drivers of fulfillment, or lack of:
  - **Maintenance / Risk:** positive performance on these attributes must be maintained to continue delighting those who are fulfilled by both the concept and product (Trial & Repeat from these consumers); else risk losing these individuals
  - **Potential / Opportunity:** improvement opportunities to attract those UN-fulfilled after product trial (Non-Repeaters)

**Maintenance / Risk:**  
% who would not be “fulfilled” if concept or product do not deliver on these



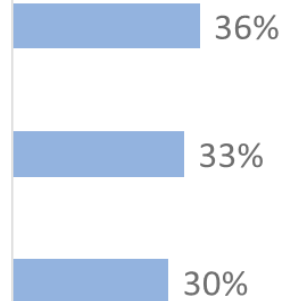
Easy to Use

Enhances Usual Routine

Lifts My Mood Overall

**Top Reasons for Fulfillment**

**Potential / Opportunity:**  
% who could be converted to “fulfilled” if perceptions on these attributes improved



Changes My Mood About the Task

Saves Time in my Routine

Makes the Task Less Stressful

**Reasons for UN-Fulfillment**



# In Conclusion

Contact Patti Wojnicz to  
continue discussion:  
[wojnp@pk-research.com](mailto:wojnp@pk-research.com)



- **Traditional hurdles may not provide strong enough of a “signal” of likely product success**
- **Additional insight and confidence in launch decisions can be gained by considering:**
  1. Performance across several contexts, e.g. blind and branded, CLT and HUT
  2. The results among consistent responders (however defined)
  3. The fulfillment or disappointment as measured in pre-use and post-use scores
    - Identifying the drivers of disappointment can be an added tool in the arsenal of identifying product improvement opportunities.