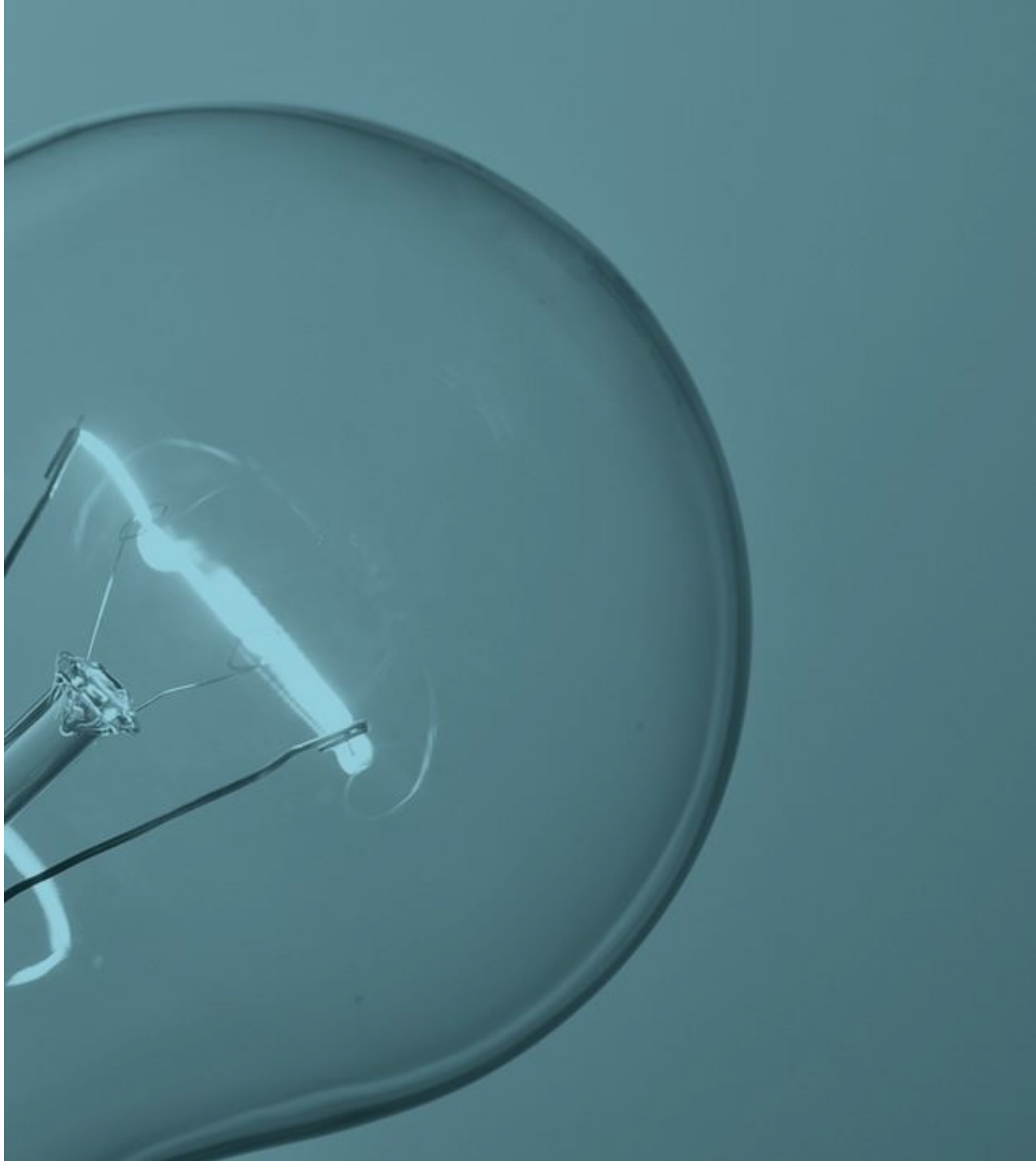
Thompson sampling in multi-attribute CBC

Isabelle Houck and Remco Don



Sawtooth Conference

5th May 2022

1. Introduction
2. Approach
3. Case study setup
4. Findings
5. Conclusions

SKIM

What brought us here?

Find the optimal product

Often, the client just wants to find out the ONE product that works best for the entire sample

A lot to be tested

How to include many different levels in one study, without exploding the model?

Interactions

Is there a combination of two attributes that work extremely well together, even though the items itself are not necessarily the best?

Use the knowledge we already have

Bandit MaxDiff* has been developed for a similar purpose: achieve greater measurement with the ability to handle large number of items. Can we do the same in a CBC setup?

Some examples...

Ice cream

Cheese

Cheese flavored ice cream

What do we want to achieve?

Methodological objective

Determine the optimal product combination while still having a robust read on attributes and their levels

Business objective

Find the best product from a (large) list of potential products, focusing mostly on the top

Use choice based conjoint to find out the optimal product from a large set of potential product combinations

Current solutions available

There are different ways to find out the best possible product configuration.

But…. There are limitations for each solution

Traditional CBC

Not adapting to individual response behavior

Limit to the number of levels to be tested for accurate read on data

Adaptive Choice Based Conjoint (ACBC)

Adapting to individual behavior, not enough information on aggregate level

Interaction effects are not considered

Synergistic Bandit Choice (SBC) for CBC

Borrowing information from sample

Focusing mainly on interactions

Bandit MaxDiff

Learning from sample behavior, while being able to handle large set of items

Focusing on one-dimensional setups

Multiple bandit MaxDiffs do not share two-way frequency balance

Our approach

Introducing: Thompson sampling in multi-attribute CBC

Choice based conjoint

Make use of the standard CBC set up for a multi attribute study

Prohibitions

Have the possibility to exclude certain product combinations that are unrealistic or outside of the scope

Sample information

Use sample information to get a more robust read on each of the tested attributes and levels

THOMPSON CBC

Test many levels

Have the possibility to test many levels per attribute, while still getting enough read across the sample

Interactions

Include the possibility of interaction effects to play a role in the product combination determination process

Find the best product

Showing only 'best' levels that are often chosen, with the main goal of finding the optimal product in the sample using aggregate analysis

SKIM

Thompson sampling in CBC

What does it do?

The algorithm utilizes an **iterative process** based on successive model estimations to be able to increase the frequency that products with high potential are shown to respondents, using the **beta distribution** to model this probability

This approach is used to solve the **multi-armed bandit problem** where the goal is to find the optimal product by constantly updating information, with the certainty increasing with every additional respondent

Advantages

- Makes use of sample information
- Has increasingly accurate estimates (*Bayesian inference*)
- Strengthens the read of the top ranked products (with potential interaction effects)
- Needs lower sample size to evaluate more levels compared to a traditional setup

SKIM

The data collection process

Counts are evaluated

Before the start of the conjoint exercise, the sample counts of each attribute level and each unique product combination are evaluated

List of product combinations determined

A selection of top levels is combined with a selection of levels with less certainty to create a final list of product combinations to be shown in the statistical design

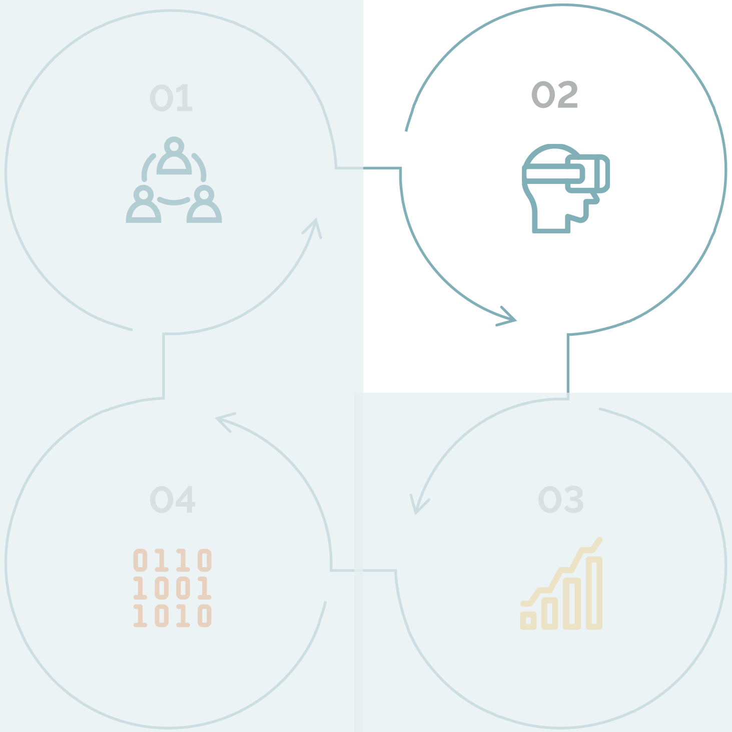
Respondents enters the survey

Respondent enters survey and goes through a series of conjoint tasks, just as in any other conjoint exercise

Counts are updated after completion

After completing the conjoint exercise, the sample counts are updated with the chosen and shown information of the respondent

How exactly?



Counts are evaluated

Before the start of the conjoint exercise, the sample counts of each attribute level and each unique product combination are evaluated

List of product combinations determined

A selection of top levels is combined with a selection of levels with less certainty to create a final list of product combinations to be shown in the statistical design

Respondents enters the survey

Respondent enters survey and goes through a series of conjoint tasks, just as in any other conjoint exercise

Counts are updated after completion

After completing the conjoint exercise, the sample counts are updated with the chosen and shown information of the respondent

Process to determine the statistical design per respondent

Determine best levels of each attribute on one-way level

Using the raw counts of the aggregate sample, rank how often **each level** is chosen versus shown **within an attribute**

Fill up with least shown levels

To give levels the opportunity to come back from a **misinformed start**, we add those that are **least shown** so far

Determine best product combinations

Rank each product combination based on the **one-way** and **two-way** counts information

Fill up with least shown product combinations

To give product combinations the opportunity to come back from a **misinformed start**, we add those that are **least shown** so far

Determine final set of product combinations to show

Determine the set of product combinations to include in the **statistical design**

SKIM

Process to determine the statistical design per respondent

Determine best levels of each attribute on one-way level

Using the raw counts of the aggregate sample, rank how often **each level** is chosen versus shown **within an attribute**

Product combinations		Attribute 2														
		1	2	**3**	4	**5**	6	7	8	**9**	10	11	12	13	**14**	15
Attribute 1	1															
	2															
	3															
	4															
	5															
	6															
	7															
	8															

Notes:

The cut-off for levels to include is maxed at 6 per attribute. Within that, 4 are determined based on the best levels (using the beta function with alpha = chosen and beta = not chosen, similar to Bandit MaxDiff).

Process to determine the statistical design per respondent

Fill up with least shown levels

To give levels the opportunity to come back from a **misinformed start**, we add those that are **least shown** so far

Attribute 2

Product combinations	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Attribute 1: 1															
2															
3															
4															
5															
6															
7															
8															

Notes:

The cut-off for levels to include is maxed at 6 per attribute. Within that, 4 are determined based on the best levels (using the beta function, similar to Bandit MaxDiff). The other two levels are added based on least shown one-way levels so far.

Process to determine the statistical design per respondent

Determine best product combinations

Rank each product combination based on the **one-way** and **two-way** counts information

Product combinations		Attribute 2														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Attribute 1	1															
	2			X		X				X					X	
	3			X		X				X					X	
	4			X		X				X					X	
	5															
	6			X		X				X					X	
	7															
	8															

Notes:

First, we look at all product combinations of the best ranked levels from both Thompson attributes.

We compute the geometric mean of the beta draw of each two way of a product combination and rank these afterwards.

With this, we fill up the first part of the final set of product combinations.

Process to determine the statistical design per respondent

Fill up with least shown product combinations

To give product combinations the opportunity to come back from a **misinformed start**, we add those that are **least shown** so far

Attribute 2

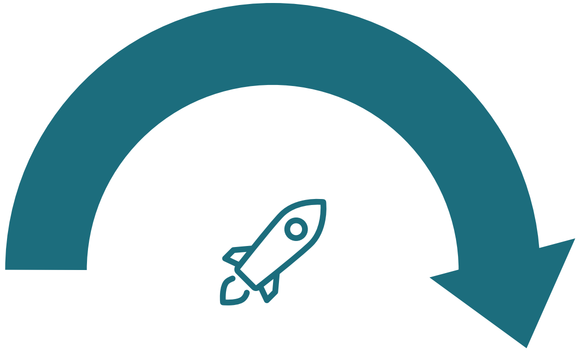
Product combinations (Attribute 1)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1			X		X				X					X	
2							X				X				
3							X				X				
4							X				X				
5															
6							X				X				
7			X		X				X					X	
8															

Notes:

Next, we fill up the rest of final set of product combinations with the ranked combinations of the top and least shown levels.

SKIM

Process to determine the statistical design per respondent



Determine final set of product combinations to show

Determine the set of product combinations to include in the **statistical design**

The final set of product combinations will be cut off at the length of *number of tasks* x *number of concepts.*

The order of this list will be randomized and will be used to fill up the statistical design of the respondent.

Notes:

- All **non-**Thompson attributes will be balanced in the way that each level will be shown an equal number of times – Standard design methods can be used for this step
- The Thompson process only starts when a minimum count is reached for each level. This gives the algorithm some information before starting the product selection process

After completing data collection, estimation is done on an aggregate level

- Using aggregate logit
- Too sparse data to use HB (same logic as Bandit MaxDiff)
- We are only interested in the optimal product
- Heterogeneity is not relevant

Case study setup

Key business question

Identify the **strongest claims** to take forward to reassure consumers on taste while also helping **establish superiority** in the category against competitor brand in the French market.

Key research questions

- What is the **best claim** to convince consumers about the great taste and naturalness of industrial soups for the client brand?
- What is the **best performing reason to believe**?
- What is the **best combination** of claim and the corresponding reason to believe?

We did a 10 minutes online test with the following modules:

Each respondents goes through one conjoint exercise

1 – Screener

Identify if respondents qualify for the survey

2a – Traditional CBC

N = 602

A series of 12 random conjoint tasks

2b – Thompson CBC

N = 599

A series of 12 conjoint tasks with an on-the-fly created statistical design

2c – Holdout sample

N = 300

A series of 12 fixed conjoint tasks

+ 1 additional tasks with the likely winning combination (based on previous research)

3 – Demographics

Final follow up questions used to get a representative sample

SKIM

Conjoint design, attributes and levels

What did it look like?

- **Attribute 1:** Brand, 2 levels
- **Attribute 2:** Main point / Claim, 8 levels
- **Attribute 3:** Reason to believe (RTB), 15 levels

Attribute 2 and **attribute 3** are treated as 'Thompson' attributes in the second conjoint. **Brand** is considered a regular attribute and its levels are balanced normally

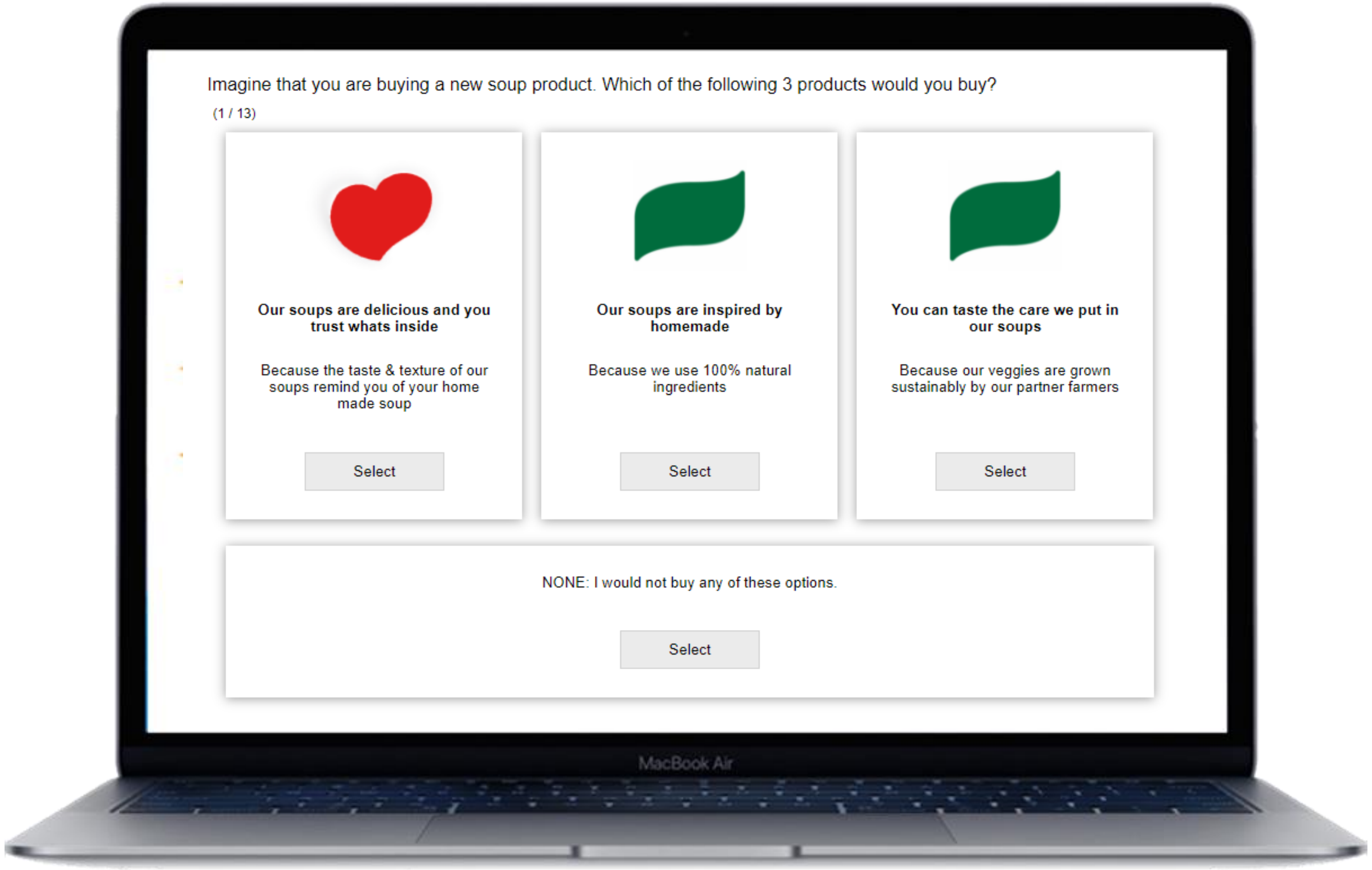
(8 x 15 = 120 potential interaction effects)

Main Point	RTB
You can taste the care we put in our soups	Because our veggies are grown with respect for the land
Our soups are inspired by homemade recipes	Because our veggies are grown sustainably by our partner farmers
Our soups are delicious, and you trust what's inside	Because our soups are cooked in France
Our soups are naturally delicious, good for you and the planet	Because our veggies are harvested in season when fully ripe
Our soups are naturally delicious and good for you	Because we use 100% natural ingredients
Our soups are naturally delicious and good for the planet	Because we don't use any artificial additives or preservatives
You can taste the difference in our soups	Because our soups are rich in vegetables
With our soups, it has never been easier to get children eat vegetables	Because we grow vegetables rich in nutrients
	Because our sustainable farming practices don't waste water or use unnecessary chemicals
	Because our vegetables are grown slowly, under the open sky
	Because we carefully select our vegetables from our partner farmers
	Because our chefs carefully select vegetables from our partner farmers
	Because our chefs have perfected the recipe
	Because the taste & texture of our soups remind you of your home-made soup
	Because our soups have a green nutriscore

SKIM

The conjoint showed three concepts next to each other

Respondents were asked to select either one of the three concepts or the none option (traditional none)



SKIM

Findings

Thompson CBC and Traditional CBC showed different optimal combinations based on the estimated aggregate logit utilities

Traditional CBC

"Our soups are naturally delicious, good for you and the planet,
because we use 100% natural ingredients"

Thompson CBC

"Our soups are inspired by homemade recipes,
because we use 100% natural ingredients"

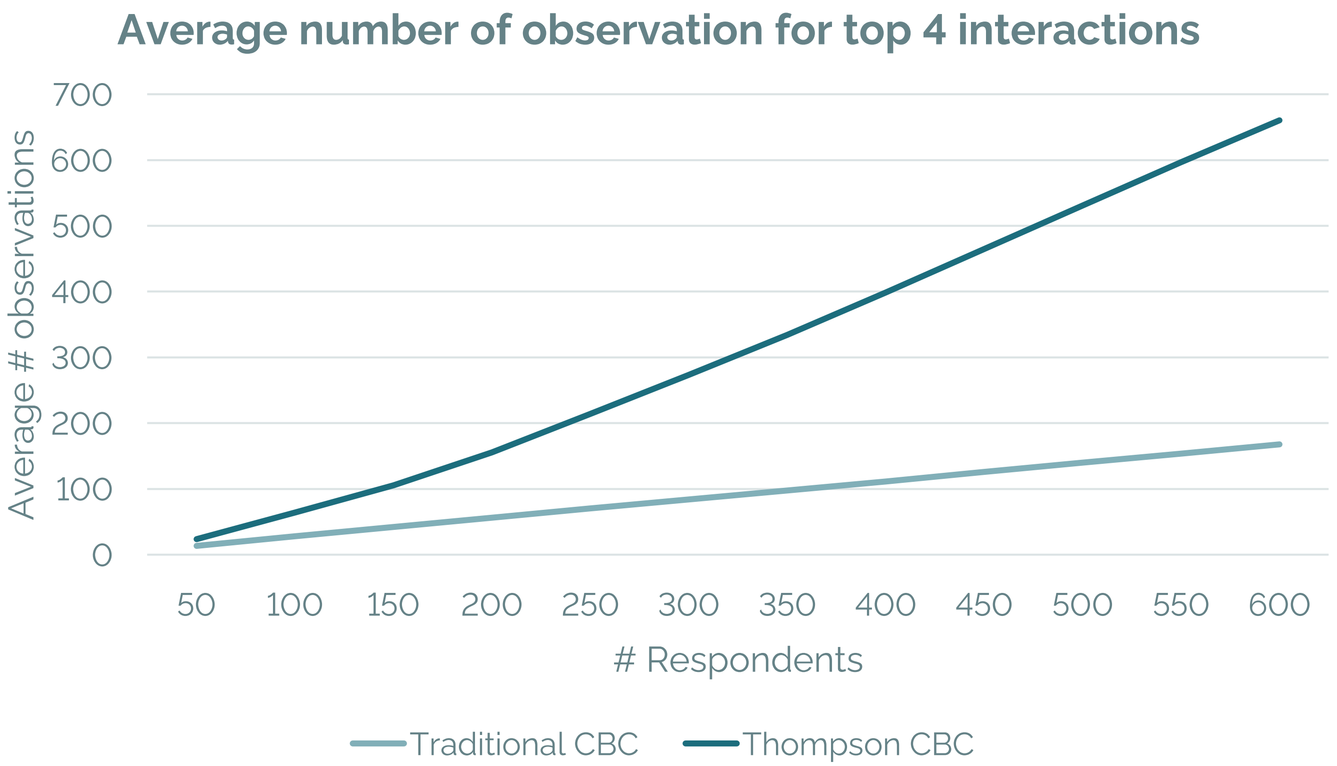
SKIM

Thompson CBC has more observations for the two potential winning combinations

Chosen / Shown	Our soups are naturally delicious, good for you and the planet*	Our soups are inspired by homemade recipes*
Traditional counts	**55%** N = 136	44% N = 228
Thompson counts	44% N = 925	**51%** N = 1140

*.... because we use 100% natural ingredients

The number of observations for top4 interactions rapidly increases over time



The average number of observations is about 4x larger, which should **reduce standard errors** during estimation.

This also means we're **showing good combinations more frequently**, which should lead to overall better predictions of the top (like Bandit MaxDiff)

SKIM

Thompson CBC has lower standard errors for the top-rated levels / interactions

Given that we showed the top levels more often in the Thompson CBC, we have a **more robust** read on these levels, both on a one-way level as on the two-way interactions between these **top levels**

The D-efficiency of the Traditional CBC still outperforms the Thompson CBC, since the overall balance is better (which is as expected)

			Traditional CBC*	Thompson CBC**
		D-efficiency	**539.0**	374.4
Standard errors	One way - ALL	**Average**	**0.049**	0.062
		Range (min - max)	*(0.037 - 0.057)*	*(0.033 - 0.089)*
	One way - Top 4	**Average**	0.047	**0.042**
		Range (min - max)	*(0.037 - 0.057)*	*(0.033 - 0.065)*
	Two way - ALL	**Average**	**0.155**	0.202
		Range (min - max)	*(0.129 - 0.188)*	*(0.085 - 0.470)*
	Two way - Top 4	**Average**	0.157	**0.116**
		Range (min - max)	*(0.136 - 0.175)*	*(0.085 - 0.195)*

** Based on 50 balanced CBC designs used in the Traditional CBC exercise with 599 respondents*
*** Based on the designs of 599 respondents that completed the Thompson exercise*

This table is created using the Test Design report in Sawtooth Software

SKIM

Predicting hold-out tasks using both Traditional and Thompson CBC shows similar patterns

Task	Concept	Brand	Claim	RTB
1	1	2	3	14
1	2	1	2	5
1	3	1	1	2
1	4	0	0	0

Hold-out Counts	Traditional CBC	Thompson CBC
25%	22%	18%
58%	52%	59%
15%	24%	21%
2%	3%	2%

	Task 1
Holdout sample share for winning concept	58%
Traditional CBC	✓
Thompson CBC	✓

Note: We should not look at hit-rates / MAE, because the purpose is **not** to predict preference shares, but rather to find the best product.

The task composition for Thompson CBC differs from Traditional CBC

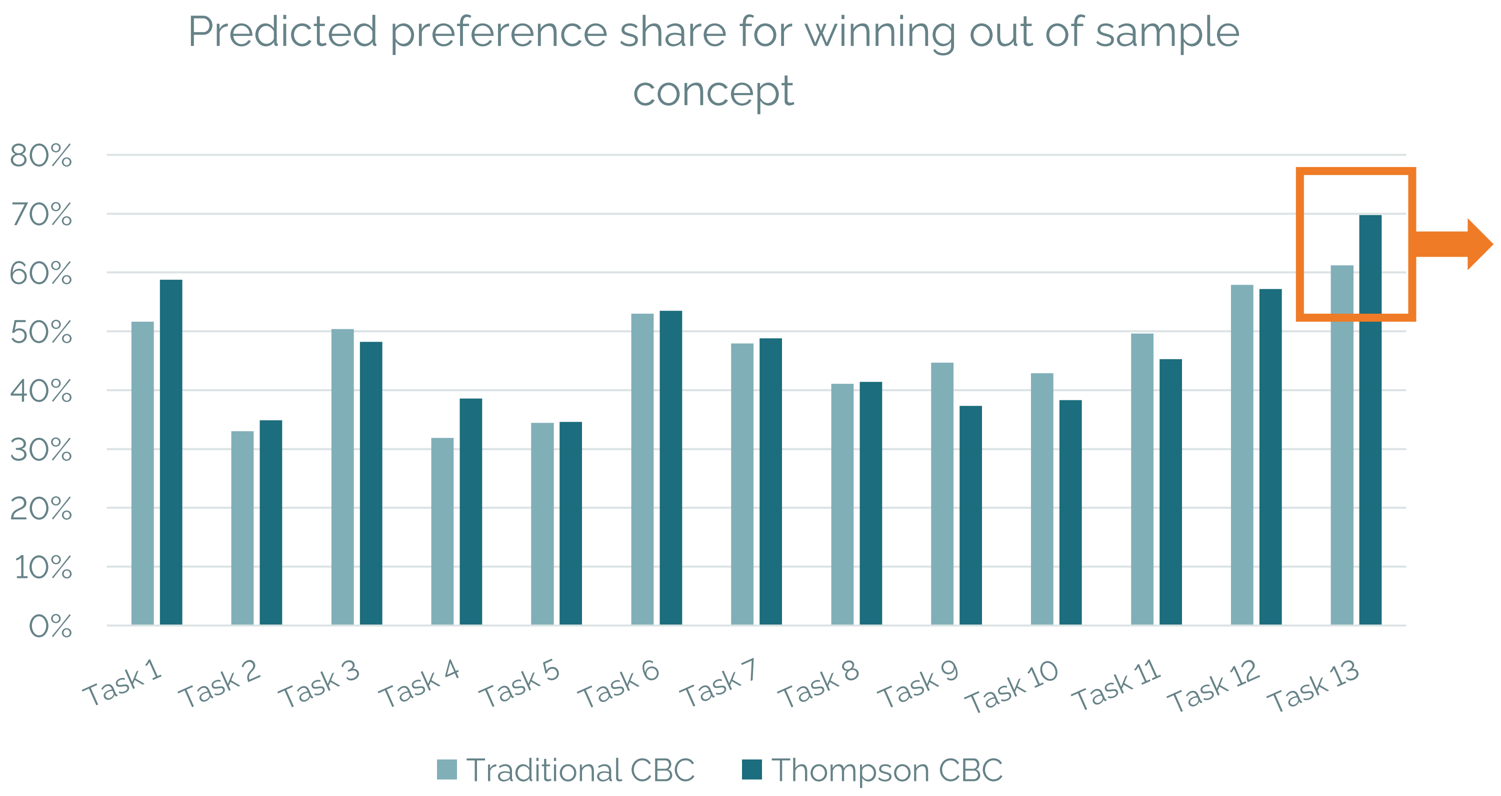
Thompson CBC isn’t showing worse predictions than Traditional CBC when predicting the most preferred concept in hold-out sample

Best concept correctly predicted per task (holdout)

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10	Task 11	Task 12	Task 13
Holdout sample share for winning concept	58%	35%	48%	37%	38%	48%	46%	46%	47%	43%	57%	54%	58%
Traditional CBC	✓		✓			✓	✓	✓	✓	✓	✓	✓	✓
Thompson CBC	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓

Tasks that the Traditional CBC incorrectly predicted only have ~35-40% share for the winning concept

Overall similar hold-out sample predictions, but higher prediction for task with highest preference share



Task 13 had the likely "best" combination based on previous research. Thompson CBC shows a **higher preference prediction** for this concept.

Likely caused by the fact that it's often shown in the optimal combination of attributes (=most preferred)

SKIM

Conclusions

Conclusions

- The results from the Traditional CBC and the Thompson CBC are **comparable**
  - This confirms that the method is valid and does what it should, given the research goal: finding the best alternative with no interest in segmentation
  - Studies with stronger interaction effects could benefit more from this approach
- Thompson CBC has **more observations** for the potential winning combinations
- Thompson CBC shows **lower standard errors for relevant interaction** effects (e.g., top combinations). Other statistical measures are more in favor of Traditional CBC
  - Since the purpose here is to find the top combination, we accept this loss in overall balance
- Thompson CBC predicts the **best concepts** with a **higher preference**
  - Since we're not using this method to predict market shares, we cannot tell whether this is good or bad

Watch-outs

- Keep in mind the research goal before using this method: looking for one **optimal product**
- Attributes to be used in Thompson sampling should be **categorical**
  - If attributes like price were to be included, you'd likely find that the lowest prices are most optimal

Thank you!

Isabelle Houck

i.houck@skimgroup.com

Remco Don

r.don@skimgroup.com