Trade-off methods: A look at the basics

Why we use them

Recommendations for using trade-off methods

Comparing strengths and weaknesses

Dr. Steven Struhl

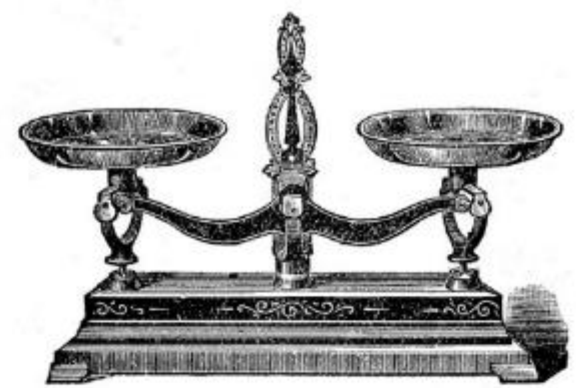
Converge Analytic

Where we will be going

Could be fun

Start

Why we use these methods
Basic assumptions



Why we use trade-off methods

- **Trade-off methods** all ask respondents to weigh specific elements or features of a product, service, claim or message against each other
  - We do this because people cannot or will not give direct ratings of individual elements that reflect what they truly value the most
- A pattern like this emerges with direct rating scales—

	How important to you is each feature?				
Features of your floor-standing wine cooler	**Not at all important**	**Not too important**	**Somewhat important**	**Very Important**	**Critical**
Lowest price	☐	☐	☐	☐	☑
Thickest insulation	☐	☐	☐	☐	☑
Genuine gold plating	☐	☐	☐	☑	☐
Built in icemaker	☐	☐	☐	☐	☑
Parking brake to prevent slippage	☐	☐	☐	☐	☑
UL listed	☐	☐	☐	☐	☑
Battery backup for power outages	☐	☐	☐	☐	☑
Extendable handle with umbrella	☐	☐	☐	☐	☑
Extra wide mag wheels	☐	☐	☐	☑	☐

- Everything becomes highly important. This is a real problem!
  - Morwitz (in *Armstrong's Principles of Forecasting*, 2002) did a very thorough review of 60+ years of research about trying to predict behavior with scaled ratings, and found no good way to use them

Broad communalities but different applications

- All trade-offs strive to uncover what is truly important in a product or service (or sometimes, a message)
- Beyond this, complexity and goals vary widely
- More complex methods more closely bridge the gap between study questions and complete, real-world decisions
- Approaches in order of increasing difficulty and complexity—
  - Q-Sort
  - MaxDiff
  - Conjoint
  - Discrete choice modeling

Does the right method wait on the other side?

- Each has applications in which it works best—choosing the right one is critical

Basic assumptions: (1) People trade off among features

- Everything in a trade-off study is a distinct feature or a distinct variation of a definite feature
- These are assumed to be measurable and comparable
  - The value of each therefore can be traded vs. other features
- This is as far into the psychology of decision-making as these methods go

The most foolish thing we could find on the psychology of decision-making

Converge Analytic

Assumptions (2): Features have distinct variations or levels that we can identify and measure

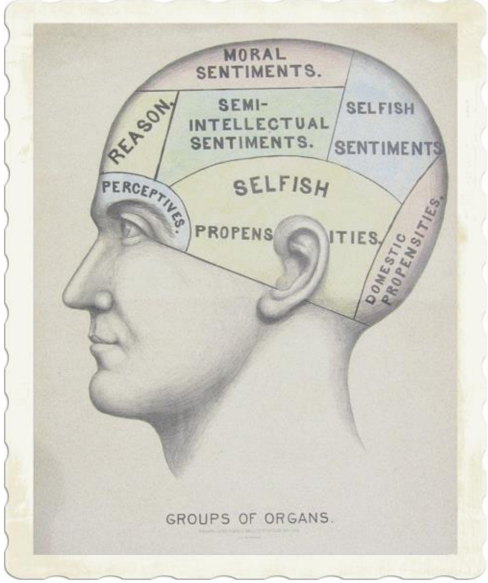
- Each distinct variation of an **attribute** is called a **level**
  - Where attributes vary continuously in the real world, they are measured only at specified points of interest in the research
    - Example: A course of treatment could be any price between $2,000 and $9,000
      - Several distinct prices are chosen to measure in this range
        - e.g.: $2,000, $4,500, $6,800 and $9,000
  - Choosing the right points to measure is critical
  - As we will see, you also need to keep to as few as possible that measure what you need to know

Choosing too much generally is not the best option

Converge Analytic

Assumptions (3) The most utility wins and utility can become share

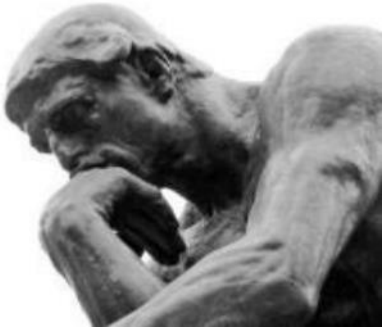
- We assume that the level of each attribute with the highest utility will win in a choice
  - This does not mean that people look at all attributes, or choose carefully
  - It **does not** means they follow a utility-based decision process
    - These methods aim to match the **outcome** of the decision process, rather than trying to decipher inner workings
  - But it does assume decisions at least are generally consistent
    - People pass this threshold in trade-off studies—if they understand them
      - Making these studies clear is critical



We are not trying to measure inner workings, just get the outcome right

Trade-off methods work best with "cognitive" features

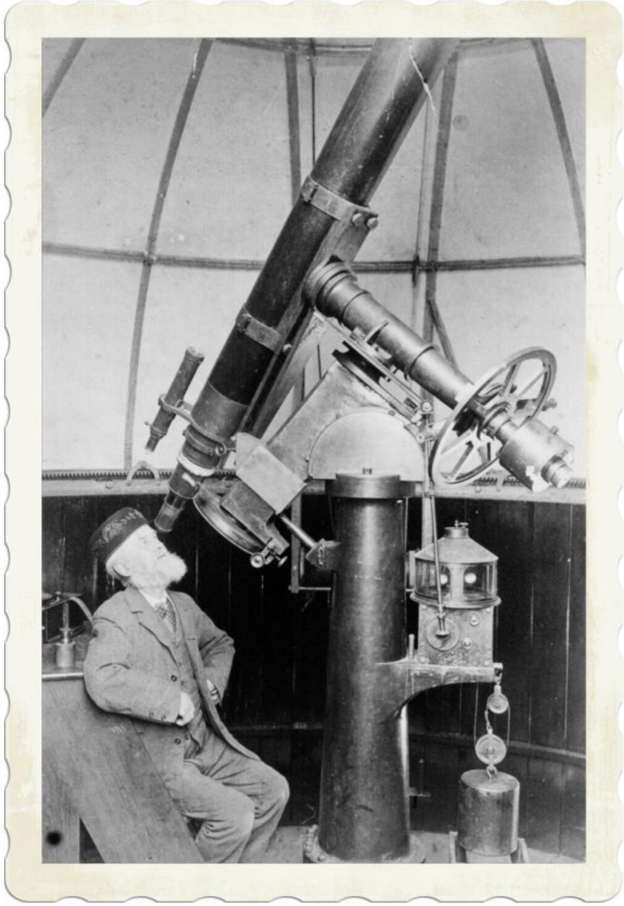
- If we consider products as ranging along a continuum—
  - From more **cognitive** (or having more to think about) to
  - More **affective** or **sensory** (or more feeling-based)
- Then trade-offs work best where products have more **cognitive** elements.



- Sometimes it is very difficult to get people to trade affective or sensory elements
  - For instance, in a trade-off exercise, people cannot accurately trade off "tastes good" against other product attributes
- However, people generally can trade off brand vs. price or other attributes
- They also can value attributes differently for different brands
  - For instance, Sony once commanded a higher price than other brands for the same set of features
    - So features once were **worth more** with the Sony name
    - This was seen in the market and in studies of choices

What the methods look like

Have you seen these?



Here is a MaxDiff trade-off: Have you seen one like this?

- A sample of one trade-off

When considering buying one of these products, which one is the <u>most</u> important and which is the <u>least</u> important?

Most Important		Least Important
○	Highest quality	○
●	Best comfort/grip	○
○	Best safety features	●

Next

- These can have 2 to 5 items compared at a time
  - Testing shows that three at a time moves most quickly
  - With three at a time, respondents do as few as 3 exercises per 4 items tested
  - So, e.g., 20 items would take 16 trade-off screens
- These responses lead to importances for the various attributes
  - Importances are **ratio scaled**, so, e.g., a score of 100 has **4 times** the importance of a score of 25

Another MaxDiff sample screen (with pictures)

- You even can trade off with pictures
- Trade-off methods can extend in many directions

Looking at these three configurations, which ONE do you like the most and which ONE do you like the least?

Like the most		Like the least
○		○
○		●
●		○

Next

Converge Analytic

Have you seen this? What a Q-sort exercise might look like

- Study participants would see a long list something like the one below
- They would be asked to pick their top 5 (or top 10 for really long lists)
- Then they would rank their three favorites
- Finally, they would do the same with their bottom 5 (or 10) and their least favorites
- That's it—the rest is in the analysis

Features you might have in your new grout cleaner: First pick your top 5	
New easy pour spigot	Buy 3 and the 4^{th} is free
Delightful pine/ozone/blackberry smell	As advertised on TV
Six pack comes with free cardboard carton	No longer sticks to clothes or hair
No rinse needed (on colored grout)	Handy travel handle
Good for camping	USDA approved
Recommended by Chef Alfonso of TV's Mighty Meals	Cleans drains too
Asbestos free	Delightful cherry flavor
Safe for pets (over 45 pounds)	Sizes over 3 gallons come with free fire shovel
New non-leaking seams (not in 64 oz. size)	Guaranteed 99% free of U 238
Turns blue when it's through	Easy open—no can opener required
Designer container (not in 6 gallon drum)	Non-GMO

Converge Analytic

And these? Conjoint cards

- Sample full-profile conjoint card
  - This one is for service delivery
  - Respondents typically see 8 to 18 of these cards
    - Online they give them ratings
    - In person, they also could sort and rank (now rare)

Feature	For this service:
Frequency of account reviews	6 months
Contract length and trial period	3 month trial period
Time on hold to reach tech support	Call back option within 5 minutes
Frequency of status updates for critical issues	Daily
Wait time for mission critical repair	Within 24 hours
Repair appointment window	AM/PM (8-12 or 12-5)
Wait time for non-mission-critical repairs	Within 4 hours
Frequency of status updates for non-critical issues	Hourly
E-mail response time	8 hours
Frequency of Status Updates	Weekly
Wait time for local telephone service	2 weeks
Wait time for high-speed internet	1 week

And finally this? Discrete choice modeling task screen

- Respondents typically evaluate 8 to 21 of these
- In each they choose the one they want, or none—or in some cases allocate across, e.g., 10 uses

Here's the first "purchase scenario" we want you to evaluate. Take a look at the different options being offered and review the characteristics of each. We realize that the features we've included may or may not be important to you. Focus on the aspects that do matter to you.

Then tell us which product, if any, you would buy.

Giant Blue	NP Hardware	Valley Computers	
Max # of users: 15	Max # of users: 75	Max # of users: 10	
Sharing data		Sharing data	
Desktop/server email			
Integrated wireless access			
Remote Access			
Work with customers		Work with customers	
Anti-virus/Anti-spam			
Storage capacity: 400 GB	Storage capacity: 200 GB	Storage capacity: 100 GB	
Storage expansion: Can add an external drive	Storage expansion: Can add an internal drive	Storage expansion: Not possible	
5 user licenses included. User license additions: $59 per user	5 user licenses included. User license additions: $489 per block of 5 users	5 user licenses included. User license additions: $99 per user	I would not buy any of the products shown here.
Price: $499	Price: $1299	Price: $299	

Definitions

Continue

The same task screen with instructions

- Instructions typically appear only on the first screen
- The definitions button is always there—keeping long explanations off the screen

Here's the first "purchase scenario" we want you to evaluate. Take a look at the different options being offered and review the characteristics of each. We realize that the features we've included may or may not be important to you. Focus on the aspects that do matter to you.

Then tell us which product, if any, you would buy.

Giant Blue	NP Hardware	Valley Computers	
Max # of users: 15	Max # of users: 75	Max # of users: 10	
Sharing data		Sharing data	
Desktop/server email			
Integrated wireless access			
Remote Access			
Work with customers		Work with customers	
Anti-virus/Anti-spam			
Storage capacity: 400 GB	Storage capacity: 200 GB	Storage capacity: 100 GB	
Storage expansion: Can add an external drive	Storage expansion: Can add an internal drive	Storage expansion: Not possible	
5 user licenses included. User license additions: $59 per user	5 user licenses included. User license additions: $489 per block of 5 users	5 user licenses included. User license additions: $99 per user	I would not buy any of the products shown here.
Price: $499	Price: $1299	Price: $299	
○	○	○	○

If you hold your mouse pointer over a term, the definition will appear.

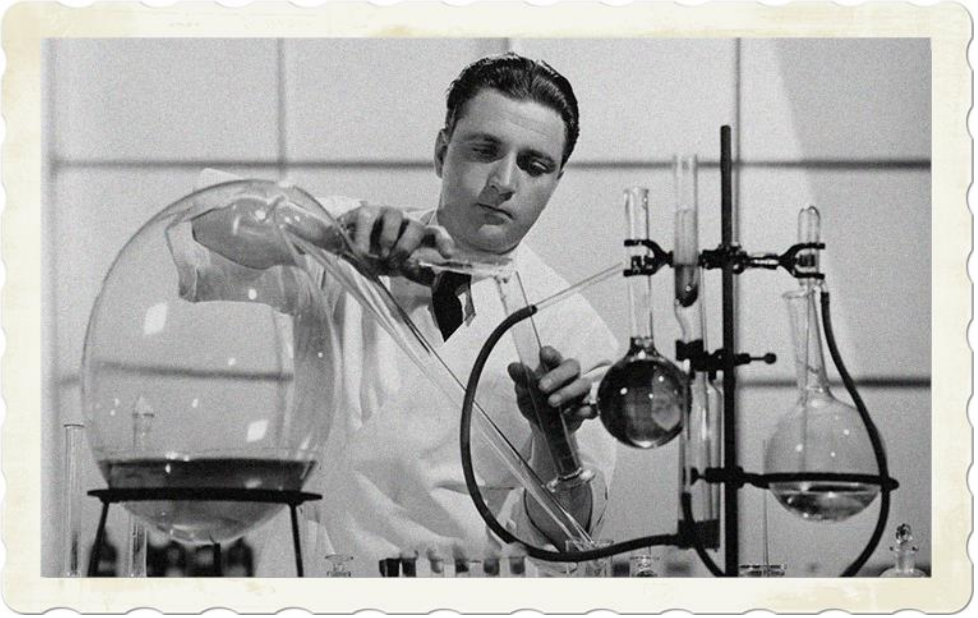
Click the button below the option you would pick.

Clicking on this box opens a pop-up window that shows all of the feature definitions.

Definitions

Continue

Determining the best uses for each method



MaxDiff and Q-Sort are limited: Features but no levels

- MaxDiff and Q-Sort provide relative importances of features
  - Note that **attribute** and **feature** mean the same thing
  - These are at the ratio level, e.g.: " Attribute A is twice as important as Attribute R"
- But they do not test multiple variations (levels) of an attribute against each other
  - For instance, we need conjoint or discrete choice if we want to know how much "shelf stable for six months" improves on "shelf stable for three months"
  - Confusing? If yes, this should get clearer very soon

Onward with all due speed

Now what is conjoint and what is discrete choice?

There has been some confusion so, first what we mean

- **Conjoint analysis**
  - A trade-off method that **looks at whole products** or (in some variants) parts of products
  - Traditionally asks for evaluations of whole products as rankings or ratings
  - Developed by market researchers
- **Discrete choice modeling** (DCM)
  - A trade-off method where we look at **whole products in the context of competitive offerings**
  - **Asks people to make a choice or choices**
  - Developed by econometricians

One of these methods will help develop the best one of these

These treat attributes differently—we will learn about these key differences later

Something of conjoint and choice: Choice-based conjoint

- A software company developed a product called **choice-based conjoint (CBC)**
  - This mixed thinking from discrete choice modeling and conjoint analysis as they then existed
- Since then the two methods, coming from different approaches, have become confusingly intermingled
  - Some have started confused
  - Some have grown confused
  - The less knowing have settled into incorrect certainty
- We of course will know everything by the time this discussion is finished

The Author in Fancy Dress as a Side of Bacon, designed by himself, which took the First Prize of Forty Guineas at the Covent Garden Fancy Dress Ball, April 1894.

Not THIS author, although it feels like this some days. This won a first prize.

Converge Analytic

Best uses for MaxDiff and Q-sort: Claims, ideas, parts of products

- MaxDiff and Q-Sort: Use these for finding importances of items that do not make a whole product/service
  - For instance, corporate claims, general concerns, basic category needs, elements in feature packages
- You get relative importances at the ratio level
  - MaxDiff provides importances for every person
  - Q-Sort does this only at the group level, but can evaluate more items

Q-Sort: What's important to all these people on average, but not just to Agnes (second over in second row)

Best uses: DCM for marketplaces, conjoint single products

- Discrete choice modeling (DCM)
  - For understanding how products or services will compete in a competitive environment, as features and prices vary
- Conjoint
  - For making the best configuration of a single product or service or service package—where competitive behavior is not important

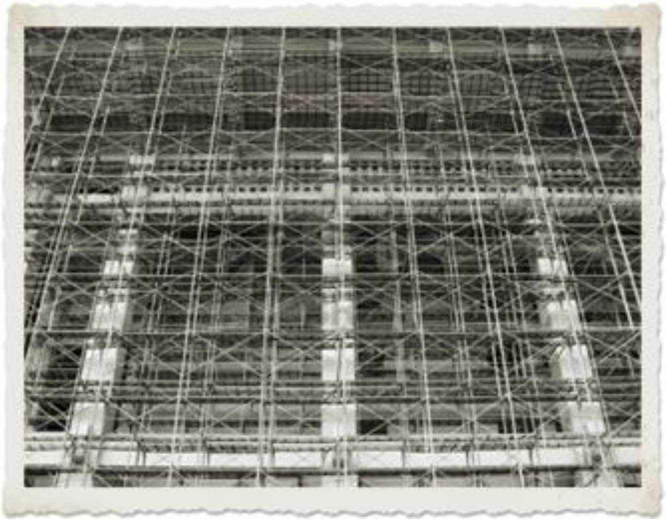
Marketplaces were hard to predict before DCM

Trade-off ground rules

Attributes and levels
Using experimental designs

The meaning of **attributes** and **levels**

- **Attributes** are a product's or service's basic features
  - In conjoint and in choice based conjoint (CBC), **brand** is considered an attribute
  - Brand can be measured for free with discrete choice modeling (DCM), and so need not be counted with attributes, as we will see
- **Levels** are **specific** variations of features that we want to measure
  - e.g., a car's fuel economy can vary from 18 to 32 mpg
    - We choose to measure at—
      - 18 mpg
      - 24 mpg and
      - 32 mpg
    - Fuel economy then has **3 levels**
  - When setting levels, the challenge is finding **the right points** to measure **without using too many**
  - Increasing attributes and levels have costs—as we will see



Not our type of levels

Thinking in attributes and levels: Interesting exercise*

- How can we express this market situation in terms of attributes and levels?
  - Four companies make **Industrial Macerators****
    - Ace (your client)
    - Hyper Size
    - Leviathan
    - Truly Big
  - These can cost between $46 and $88 million
    - Ace, however, considers itself the quality leader, and will not stoop to any price less than $52 million
  - They have some very special features, namely—
    - 2, 4, or 6 macerating paddles
      - Ace has just patented an 8-paddle design, which it wants to introduce
    - 3 to 17 sparging poles
    - A wide range of colors: black, brown, olive drab, and pink

Something like this only much bigger

** Not really a quiz*
*** Don't worry; this one is not a real product—at least we hope it isn't*

Maceraters in attributes and levels: Can you answer?

- First consider the attributes as very well-defined, specific features--things you can point to or show. What would you include?

- Now consider these attributes in terms of benefits or functions useful to the user. How would you describe them?

Macerators in attributes and levels: Sample responses

First consider the attributes as very well-defined, specific features--things you can point to or show. What would you include?

1. Prices from $46 million to $88 million. Make sure you include $52 million. Maybe more if they want to raise prices. So maybe $46, $52, $60, $72, $88 and $94

Comment: This is a good try. But it is a lot of prices. As we will see, prices can be specific to brands

2. Number of macerating paddles. Maybe 2, and 4 and 8 to test the new patented design

Comment: Good job. This would work!

3. Number of sparging poles, say, 3, 5, 7, 9, 11, 13, and 17

Comment: This is a lot of levels. We should restrict how many we test based on our in-depth understanding of the pole market, measuring only what is important

4. Brand: The four brands

Comment: As we will see, brand does NOT count an attribute, but something more

Comment: And do not forget color as an attribute—the marketing team would be heartbroken

Meaning and macerators: An important lesson about knowing the market

- Now consider these attributes in terms of benefits or functions useful to the user. How would you describe them?

Now you have me. You would have to have some idea of how these things actually worked and what they did.

Comment: This is the truth about all trade-off studies

We really need to understand the market before we start them

It's not certain that this has much to do with really knowing what you are doing. Your author just really likes the image

Converge Analytic

Just to underline the message: Know the category first

- We just saw that you need to know the category to make good decisions about which attributes and levels to test
- If your knowledge of the category is scanty, you likely need some qualitative research first
- Your knowledge will help make the exercise more compact as well
  - As we will see later—
    - You need to be sparing with the attributes and levels you test
    - The more you test, the more screens or tasks your study participant has to go through, to get you enough information

Maybe being a total miser is too much, but you do need to be economical

Focus on benefits, not internal workings

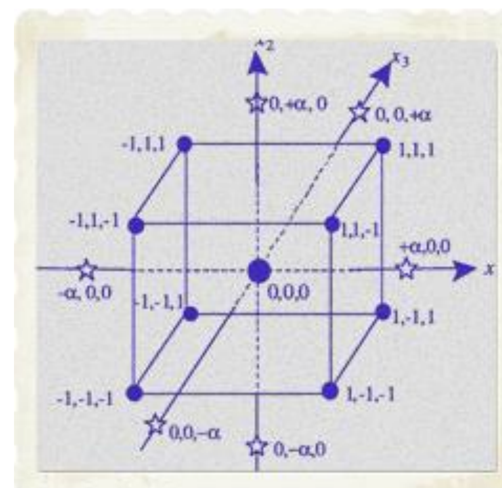
- It is easy to get fixated on product descriptions, not benefits
  - Customers typically care about what the product can do for them, not how it is put together
    - Clients who make products live with them all day, and so small details mean a lot to them
    - This is good for product quality, but not for testing responses
  - Our job often includes moving the focus to where it belongs: the product's users, and what they see and want in the product

"I want the time, not how the watch is made"

Converge Analytic

Experimental designs give conjoint and choice great power

- **Experimental design** covers a broad range of approaches
  - However, all designs for trade-offs meet one goal
    - Accurate estimation of many different situations using relatively few carefully selected situations or comparisons
  - That is, if we use an experimental design and show just a few **stimulus items** (products, marketplaces, or comparisons)
    - Then we can estimate accurately what would happen in hundreds, or even thousands, of different situations
- Confusing? Let's go on to more about what these are. . .

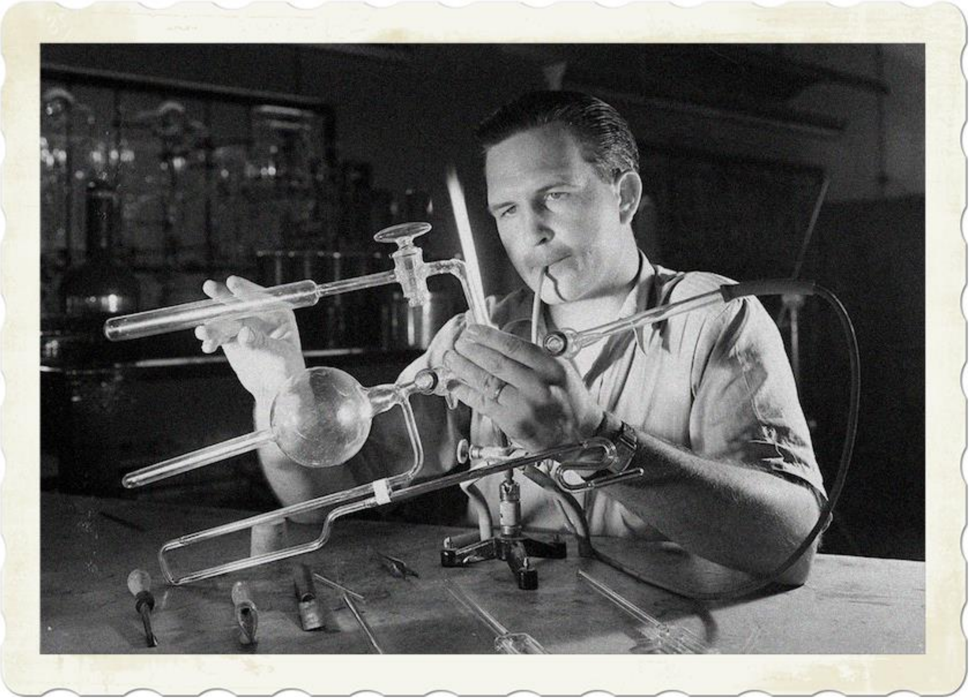


This should clear up everything!

Converge Analytic

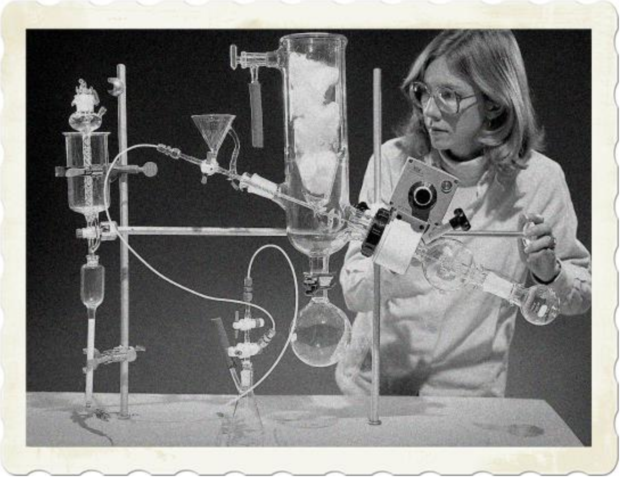
What on earth is an experimental design?

And why do we use them?



What is an experimental design?

- There are many types of experimental designs, but again all of them have the same aims
  - To measure the effects of changing one factor or several factors on some outcome
  - Accurate estimation of many different situations using relatively few carefully selected situations or comparisons
- Good experiments make sure that we are measuring accurately whatever we want to quantify
  - Good experiments observe the rules that guarantee things work
  - There are infinitely many bad experiments, and we need to avoid those



This really will not work out without planning

Converge Analytic

Benefits of experimental designs

- These designs allow us to measure the effects of more than one feature varying at the same time without the effects getting tangled
  - Suppose you cared about responses to changes in a car's horsepower, fuel economy and acceleration
    - You would need to make a number of measurements of how people responded, in which each factor varied, and see the differences
    - You would want pure measurements—no contamination from the effects of one factor varying mixed up with effects of another factor varying
- Designed experiments are hard to devise
  - A lot of ingenious and persistent people worked very hard to get these designs just right

Experimental designs took off on the farm, trying to get better crops—it's a long story Maybe another time

Setting up an experiment: attributes and their levels

- Here, we are asking people to rate how likely they would be to buy each of several cars, based on horsepower, mileage, and time from 0 to 60
- We would have no idea what is influencing preference using the bad setup to the right—everything improves in the same way from car 1 to car 3
- We get a clear picture from the unrelated variations on the left

BAD

How likely would you be to buy each car?

	HP	MPG	Seconds 0 to 60
Car 1	120	30	9
Car 2	150	40	8
Car 3	180	50	7

GOOD

How likely would you be to buy each car?

	HP	MPG	Seconds 0 to 60
Car 1	180	30	9
Car 2	120	40	9
Car 3	240	50	9
Car 4	240	30	8
Car 5	180	40	8
Car 6	120	50	8
Car 7	120	30	7
Car 8	240	40	7
Car 9	180	50	7

- We also need to show more than three hypothetical cars to measure what influences responses . . .in just a bit, we will get to determining that number

The good design has absolutely no measurable relationship in the way features vary

- We use standard **correlations** to measure the relationships that variations in the factors or features have to each other
  - That is, we consider each car a **row** in a grid of numbers and
  - Each feature (or factor) a **column**
- Each feature therefore is a **variable**—the same way columns are variables throughout statistics
- And indeed, in our example, all correlations are zero
  - That is, there is no relationship in that they vary from one hypothetical car to the next
    - There are blanks in the table because variables cannot have correlations with themselves

Correlations			
	HP	MPG	0 to 60
HP		0	0
MPG	0		0
0 to 60	0	0	

No relationships

Converge Analytic

Designing: The computer is still doing more work

- Another feature of the design
  - Every **pair** of attribute levels will appear at least once
    - This is **not** every three-way set, four-way set (or more)
  - With the design we just did, you will see, e.g.:
    - 120 HP with 30, 40 and 50 MPG each at least once
    - Same for 150 HP and 180 HP—each appears with each MPG at least once
    - 120 HP appears at least once with 7, 8, and 9 seconds to 60
    - And the same for 150 HP and 180 HP—and so on
  - Getting everything right takes a lot of work
  - Fortunately, any relatively new computer can easily crank out these designs

One of these is not required

The great power of experimental designs in measuring many things

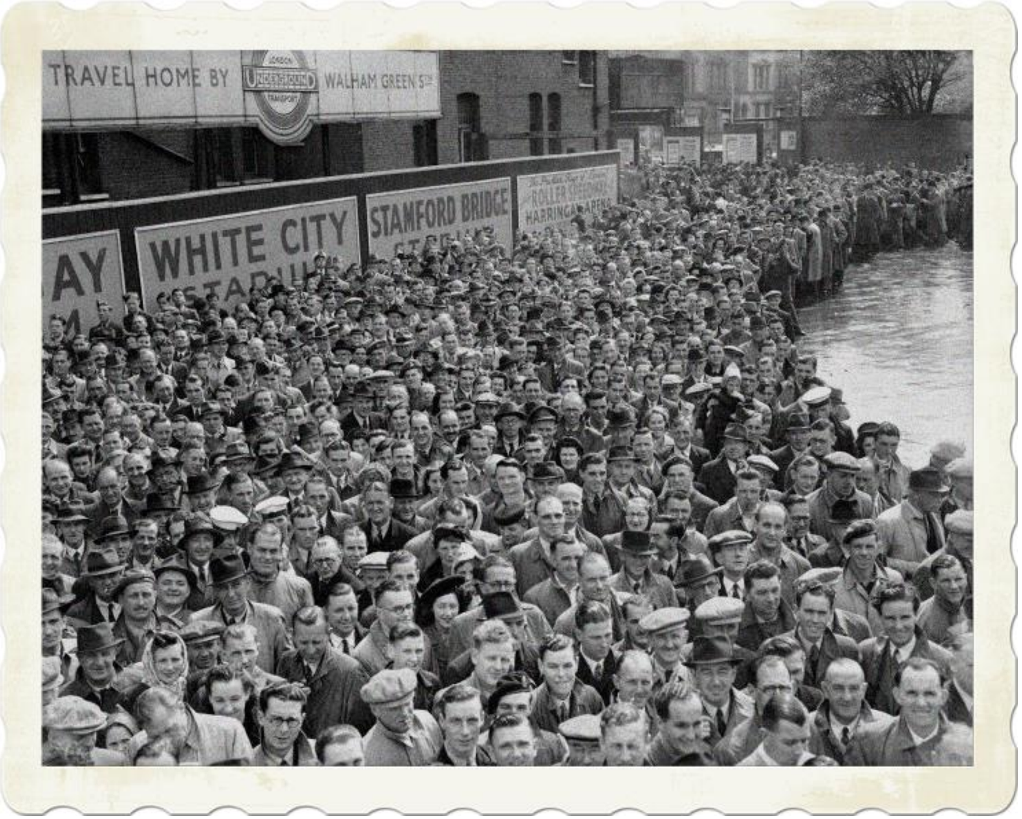
- Finally, some examples showing why we bother
- Suppose you had a product with—
  - 6 attributes, each having 3 levels, and
  - One attribute with 6 levels
- This would mean that you could have 3 x 3 x 3 x 3 x 3 x 3 x 6 or some **4374** possible variations on this product
  - Using an experimental design, we can accurately estimate the value all **4374** possible variations using **only 18 product descriptions**

Practically rocket science

- Suppose you have a product with 18 two-level attributes
  - This would give you 2^{18} (2 to the 18th power) or **262,144** combinations
  - You can measure all these possible combinations **using only 20 product descriptions**

Converge Analytic

How many screens and how many people?



Just how big is that experiment?

- Our example was an actual experiment
  - Let's lay it out again . . .

Three attributes
Each one is a **variable**
We read how each changes
by looking down each column

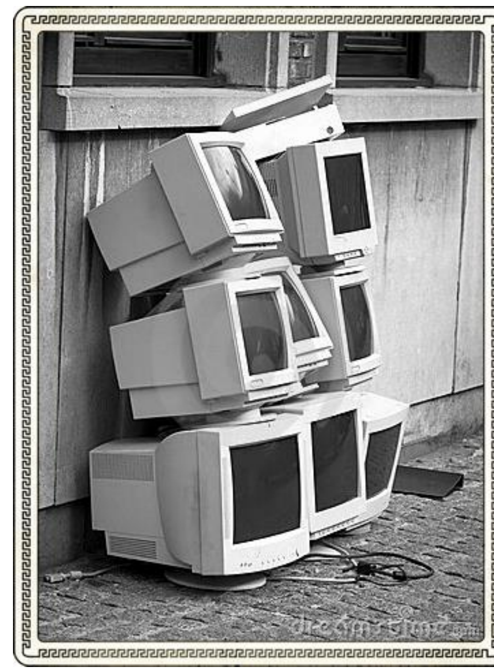
Nine rows
Each one is a **screen**
a person would see online.
Some still call these **cards**
or **boards**, going back to
the days before online testing.
Some call these **tasks.**

	HP	MPG	Seconds 0 to 60
Car 1	180	30	9
Car 2	120	40	9
Car 3	240	50	9
Car 4	240	30	8
Car 5	180	40	8
Car 6	120	50	8
Car 7	120	30	7
Car 8	240	40	7
Car 9	180	50	7

- But why nine screens (or cards or tasks)?

Sizing the experiment: How much we need to show

- We count the numbers of attributes and levels measured to understand the size of the experiment
- We need to show more stimulus items (more screens with hypothetical products) as we measure more
- This is a rough rule of thumb—the exact formula is a little more complicated
  - 2-level attribute: 1 row or screen shown
  - 3-level attribute: 2 rows or screens shown
  - 4-level attribute: 3 rows or screens shown
  - 5-level attribute: 4 rows or screens shown
  - 6-level attribute: 5 rows or screens shown
- And then we need to consider a bit more. . .

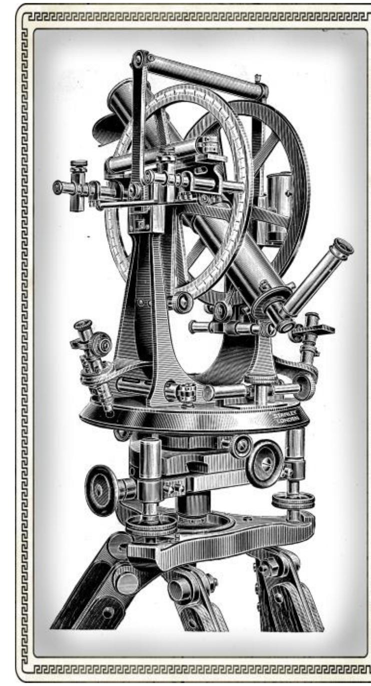


Eight screens,
but not what we mean

Converge Analytic

Finishing our count for the size of the design

- Then we add two more screens
  - One for measuring the **error** in the model
    - This allows us to know how well we are measuring
  - One for a very useful term called the **constant**
    - This has mathematical meaning, but we use it to measure **the value of the brand or the choice**
- One last wrinkle
  - The design must be at least as big as the **product or the two largest attributes**
- In our example
  - Three 3-level attributes = 3 x 2 or 6 screens + our two extra for measuring error and the constant = 8
    - However, we must have at least **3 x 3** (the product of the two largest attributes) or 9 **screens**



We need a means to determine accuracy

When people get tired of doing trade-off exercises: quickly

- Making choices in a survey wears out study participants
- Some studies show people can do well with up to 21 screens
  - This may work with highly interested and/or well compensated participants
- Most adults can handle **12 to 16** reasonably well (with grumbling)
- With less literate, children and the uninvolved**, about 10** is the limit
  - But we use up 10 screens with only four 3-level attributes
  - We use up 15 screens with five 4-level attributes
- That's not much measurement
- We usually want to do more
- What to do?

Illegal stimulants not recommended

Converge Analytic

The old remedy: Split the design, add more people

- We used to split the design up, giving a fraction to each person, adding more people
- For instance, a client goes crazy and wants this
  - Six 4-level attributes, six 3-level attributes and twelve 2-level attributes
    - We would need 48 screens
- Old solution
  - Show each person 12 of the 48 and **multiply the number of study participants by 4**
  - With choice models, this worked!
  - But it made for much bigger and costlier studies
- Splitting up designs for conjoint was very messy and likely to explode

Nice idea, but still one person

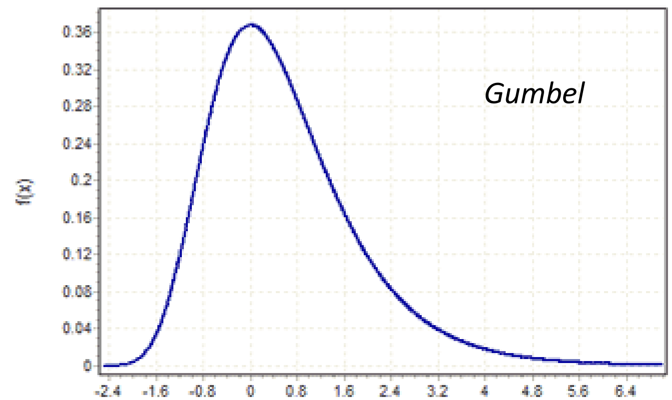
Hierarchical Bayesian (HB) analysis to the rescue

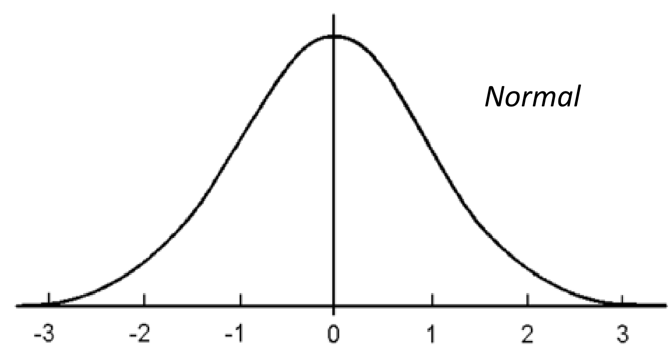
- **Hierarchical Bayesian** (HB) analysis really stretches how much we can measure with trade-offs, but relies on some fairly mind-boggling concepts
  - It has been proven under fire—since the 1990s
- With HB, we can measure **3 to 4 times as many attributes** in choice models/conjoint
  - We could, e.g., reasonably split a 48-screen task into 12-task sets, show each person one 12-task set and not increase the sample
  - As a big bonus
    - **Individual level data** from choice models (and MaxDiff)
    - Never possible before HB
      - Those were the bad old days

HB analysis makes other methods seem old and tired

So, how many people in total?

- First, no hard and fast rules exist
- Experience shows that, for a reasonably sized experiment, **125 per group** you want to measure separately is safe and reliable
- Some say 200, very cautiously
- Samples work harder with DCM
- More technically—
  - Error for discrete choice follows a non-normal (Gumbel) distribution
  - It is somewhat tighter than a normal distribution (bottom)
  - Therefore DCM samples act like bigger samples with smaller errors
- You may even get away with somewhat fewer than 125 with very homogenous groups





We lied before.
This is our most boring illustration.

The toll in larger samples of monster experiments

- With very large experiments, varying many attributes and levels, you may still need to increase the sample to get enough for measuring
  - This likely happens when you need more than 48 screens
    - Some experts would say this is stretching too much—and might say any design requiring over 36 screens must have an increased sample
- One example—
  - A study showed 77 candies** on a simulated shelf
    - 55 of them could be there or not
    - The design for this took 60 screens
    - Each person saw 10 out of the 60 screens
  - We used a sample of 1000, and allowed no more than 2-way subsamples
    - 500 was the minimum group size analyzed
  - By the way, predictions were accurate

***This only happened because of a totally out of control situation. It was a small fraction of the 400+ that the client wanted to test*

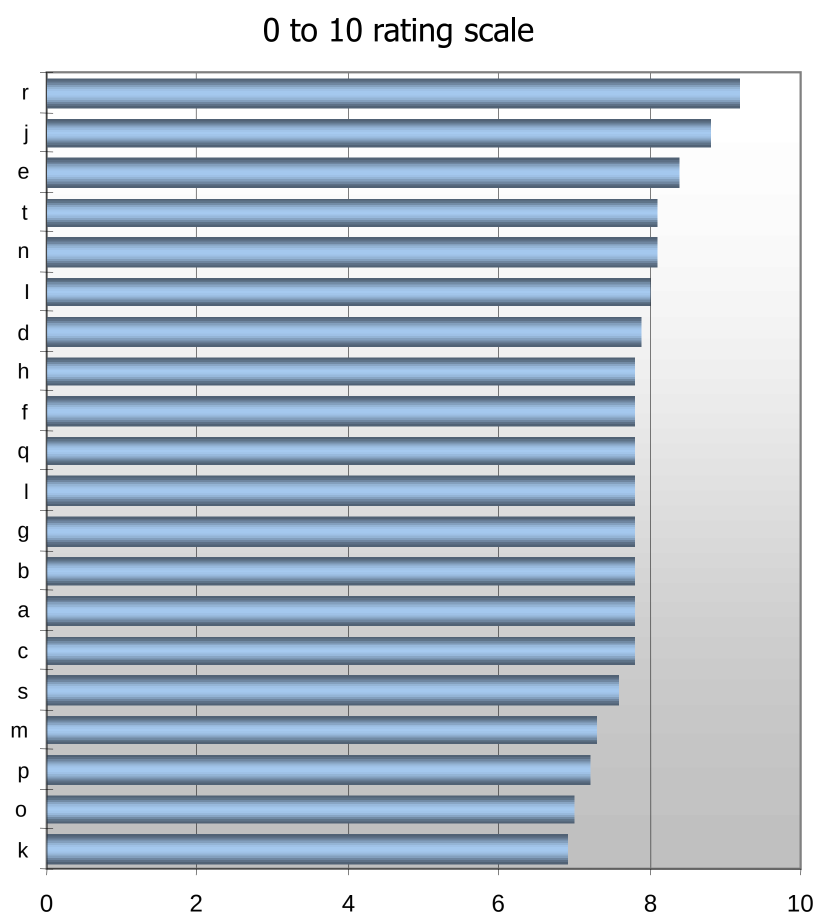
There have been a lot of candies for decades now

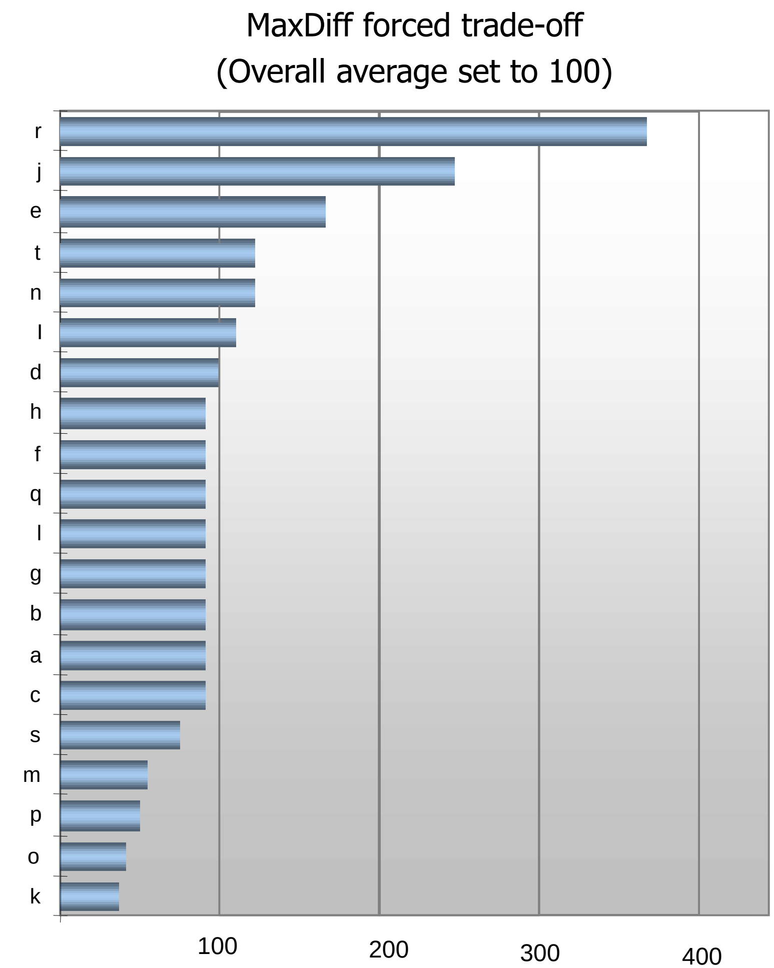
Converge Analytic

Helpful output you can expect

Helpful output from MaxDiff: A real fix on importances

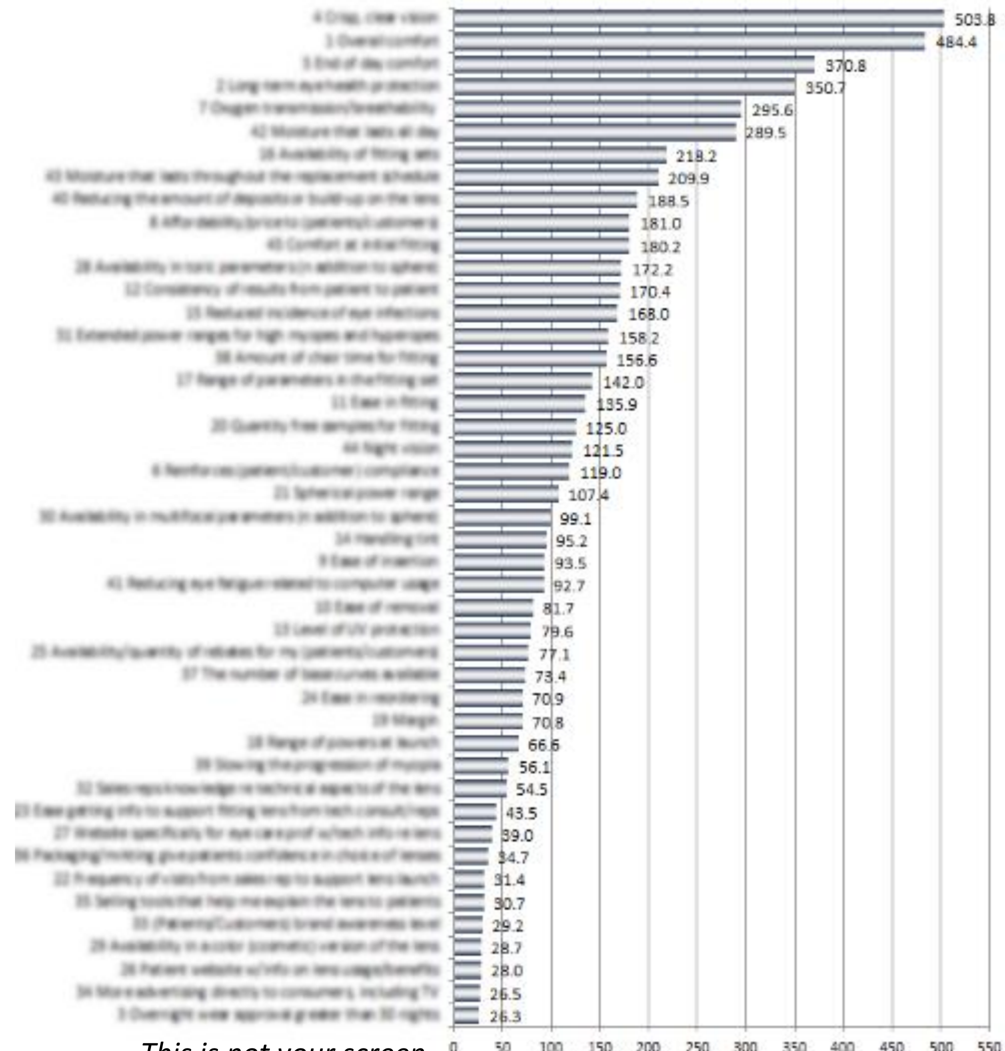
Here the same attributes in were tested in two ways: standard 0 to 10 rating scales and MaxDiff. MaxDiff shows differences much more clearly. MaxDiff is indexed so average = 100. The top item at about index ≅ 380 is about nine times as important as the bottom at index ≅ 40.





Helpful output from Q-sort: Large numbers of items prioritized

- A disguised list of about 55 items from a recent study
- List is indexed so average importance = 100
  - Two clear winners are about 5.0 and 4.8 times as important as the average
    - Index values 503 and 484
- Lowest items index at 26.3 and 26.5
- The top item is about 20 times as important as the least



This is not your screen.
It is the disguise section.

Converge Analytic

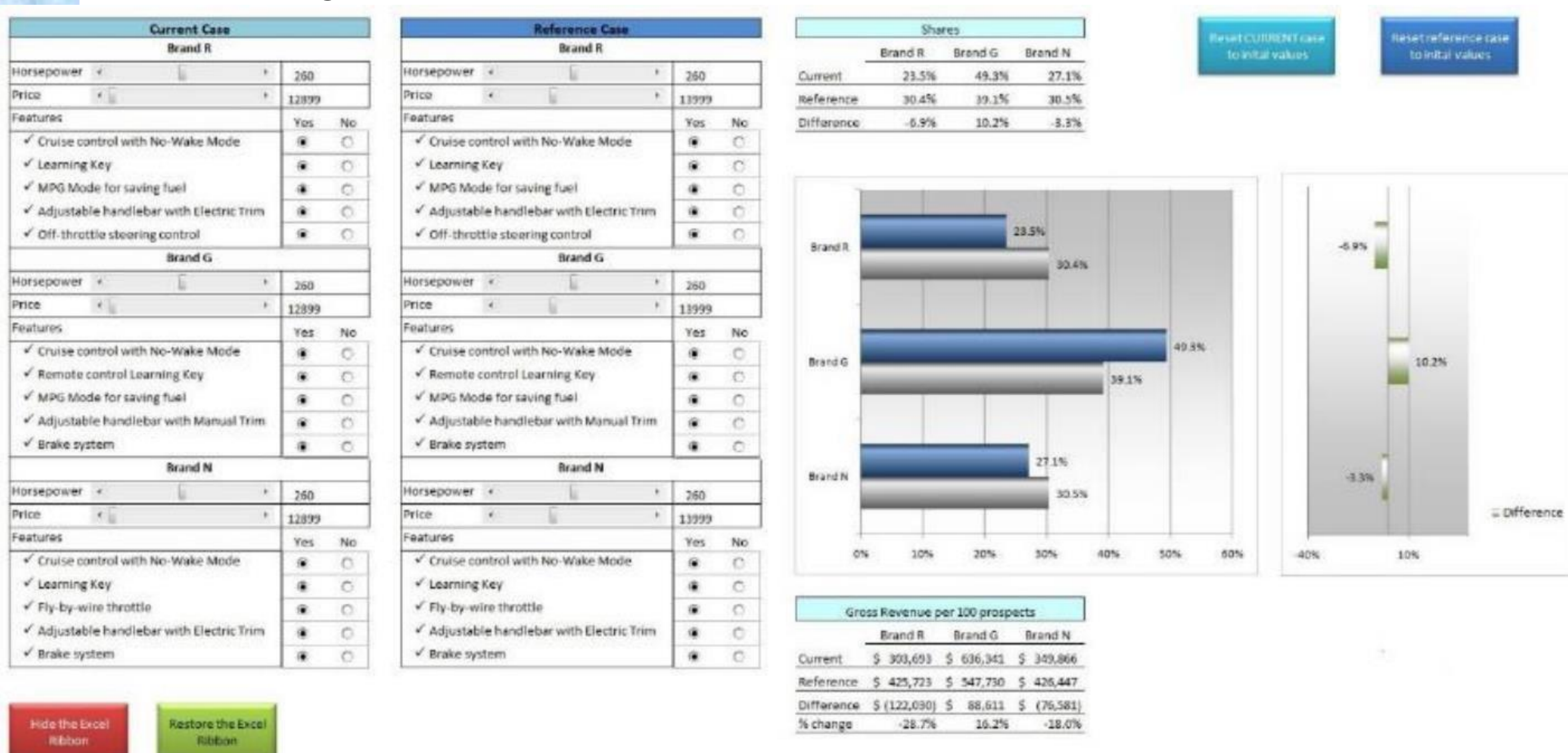
What can you expect from DCM and conjoint?

- First and foremost, and most helpfully, a market simulator
  - Typically runs under Excel and allows you to test all possible combinations in real time
  - These are run with easy-to-use controls
- Also, specific simulations statically, in a presentation
  - These would show the results of setting specific product configurations
- Possibly, these could compare each brand's response to changes in price, again statically in a chart
  - A brand's changes in response to changing its prices = **self effects**
  - Effects on other brands from changing one brand's prices = **cross effects**

Not guaranteed to produce helpful output

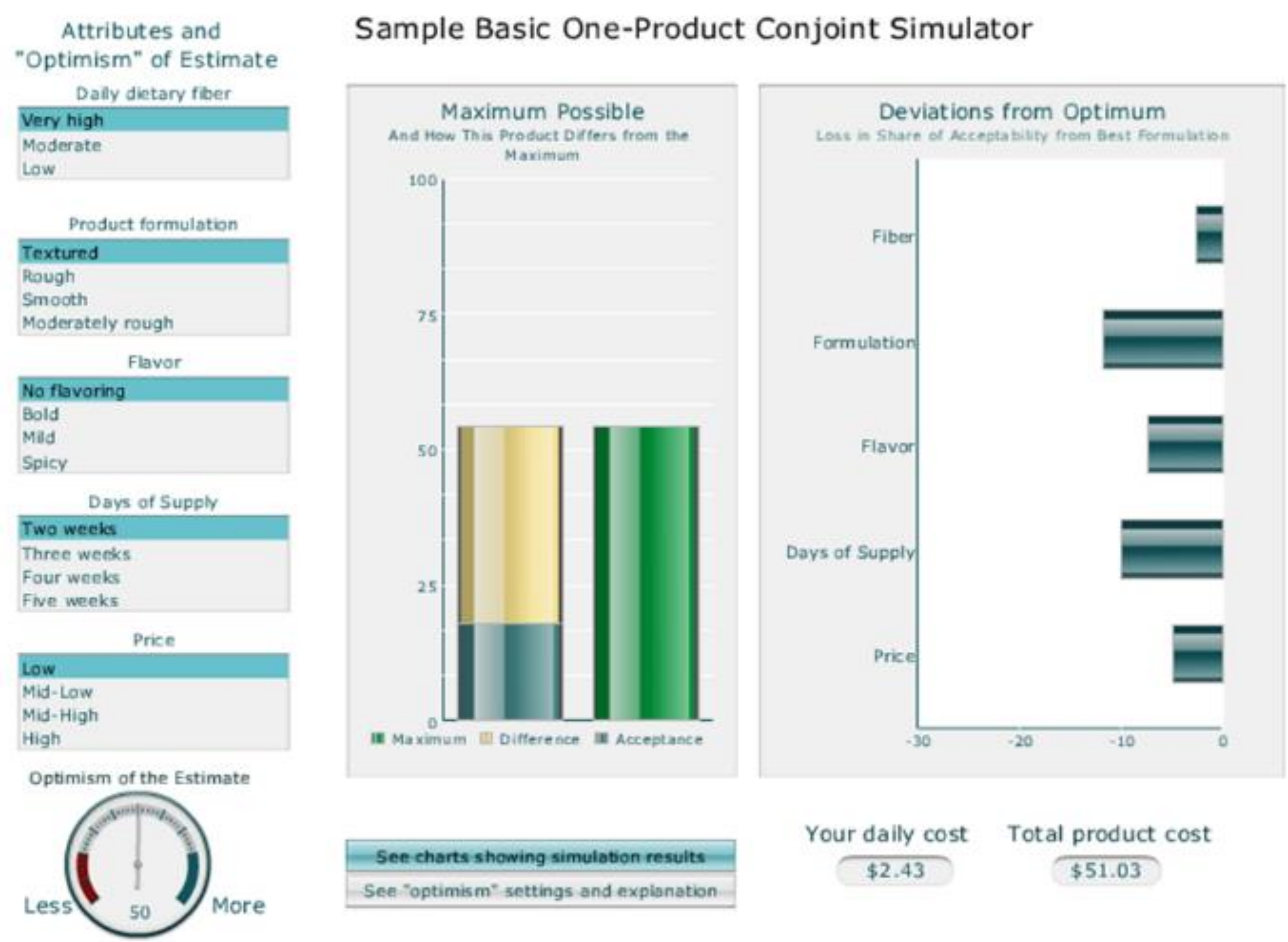
Some helpful output: Market simulator programs

- These easy to use, Excel—based programs give real time answers to hundreds or thousands of “what if” questions about varying prices and features
  - They also provide both graphical and numeric displays of results and have controls (drop-downs, sliders, etc.) to simplify use
  - Results stay up front in their most useful form and calculations remain hidden where they belong



A section of a demonstration simulator

A picture of an interactive simulator that is located at www.convergeanalytic.com/images/PDFs/samplesimulator.swf. Look for the link!

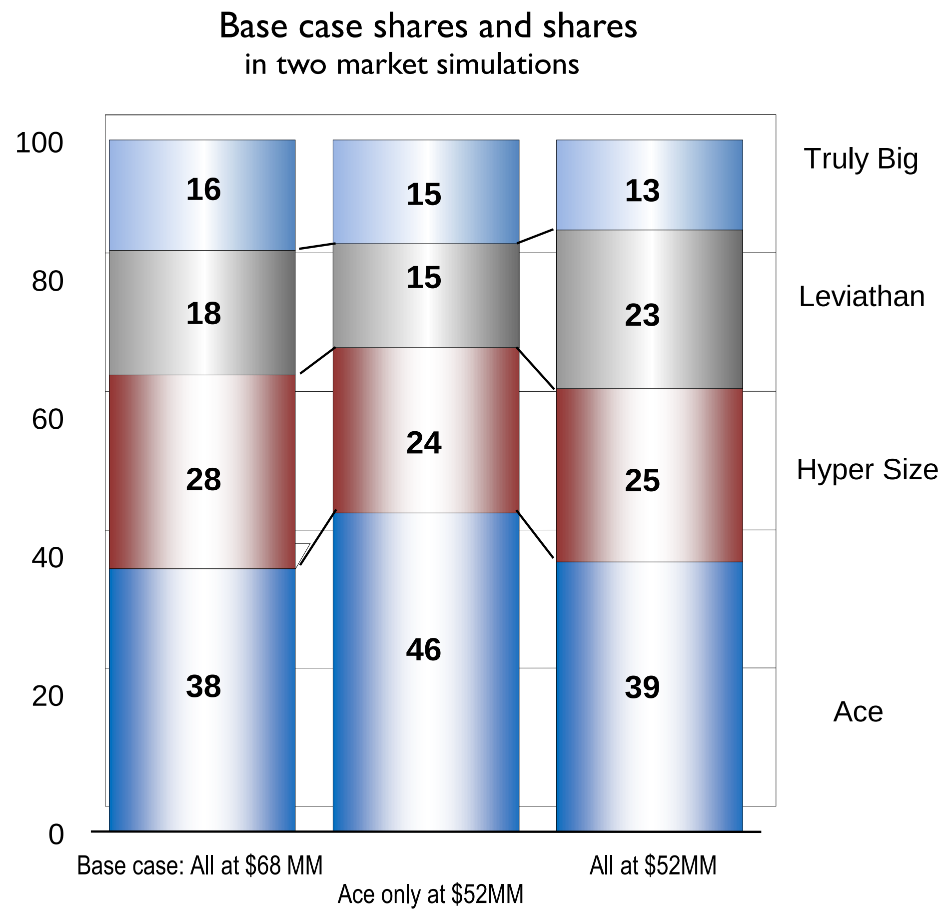


The original can run inside PowerPoint as well as on the Web

Converge Analytic

Some helpful DCM charts: Changes in different scenarios

- Here, how all shares change in two different competitive scenarios, compared with the reference or base case
- A very dramatic way to show answers to a key "what if" question
  - Insights gained from this analysis and display often **make audiences' eyes light up**,[1] and indeed can repay all your hard work
- These are only a few of the types of displays that can flow from a choice-based modeling analysis
- **Quick quiz**: What is the crucial lesson for Ace[2] from these two simulations?
- **Quick answer**: Do **not** start a price war and hope that nobody else does
  - Leviathan is the only possible winner if this happens: share up 5 points on a base of 18, or 28%, while price per unit decreases 24%



[1] Or cause calls to lock up the results, so they can't leak out to competitors

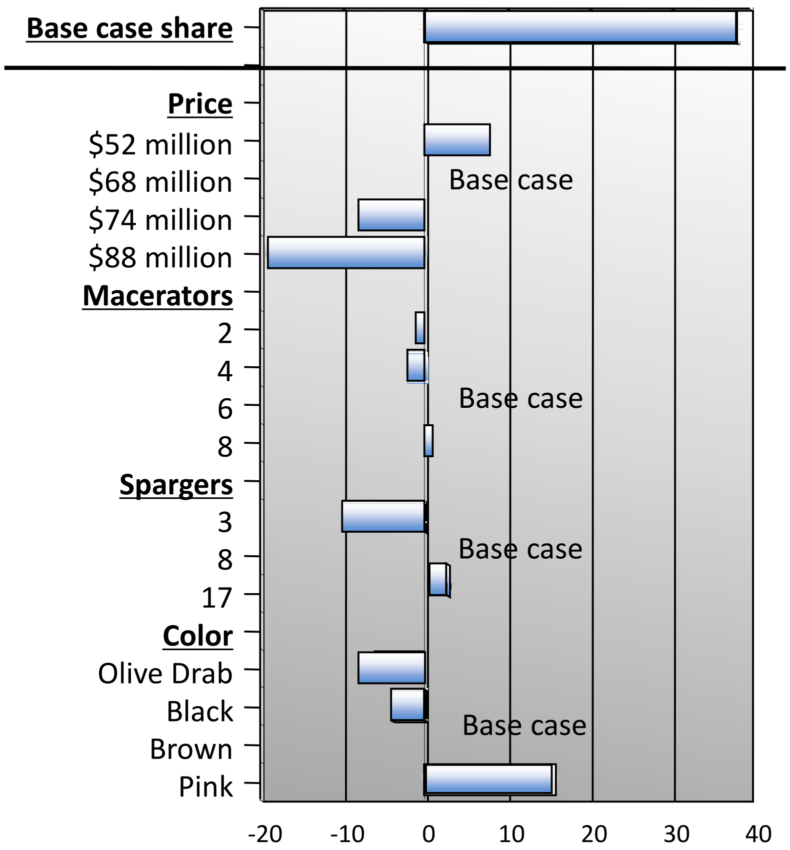
[2] Remember them, all the way back on slide 27?

Converge Analytic

Some helpful DCM charts: Showing feature changes one at a time

- This chart gives a quick overview of the relative effects of changing attributes **one level at a time** for a brand
- Here is a report for the Ace Enterprise Macerator
- It shows what happens when Ace varies but all other brands are held at set values (their **base case or reference case**)
    - All attributes for Ace are varied one level at a time
    - Results are saved
  - When the next brand (Leviathan) varies, Ace and all others stay at the same set values (their **base cases**)
    - This repeats for all brands
  - This one chart reflects the results of 15 simulator runs
- Note that **base case for Ace** always appears as **zero deviation** in the chart:
  - Price: $68 million
  - Macerators: 6
  - Spargers: 8
  - Color: Brown

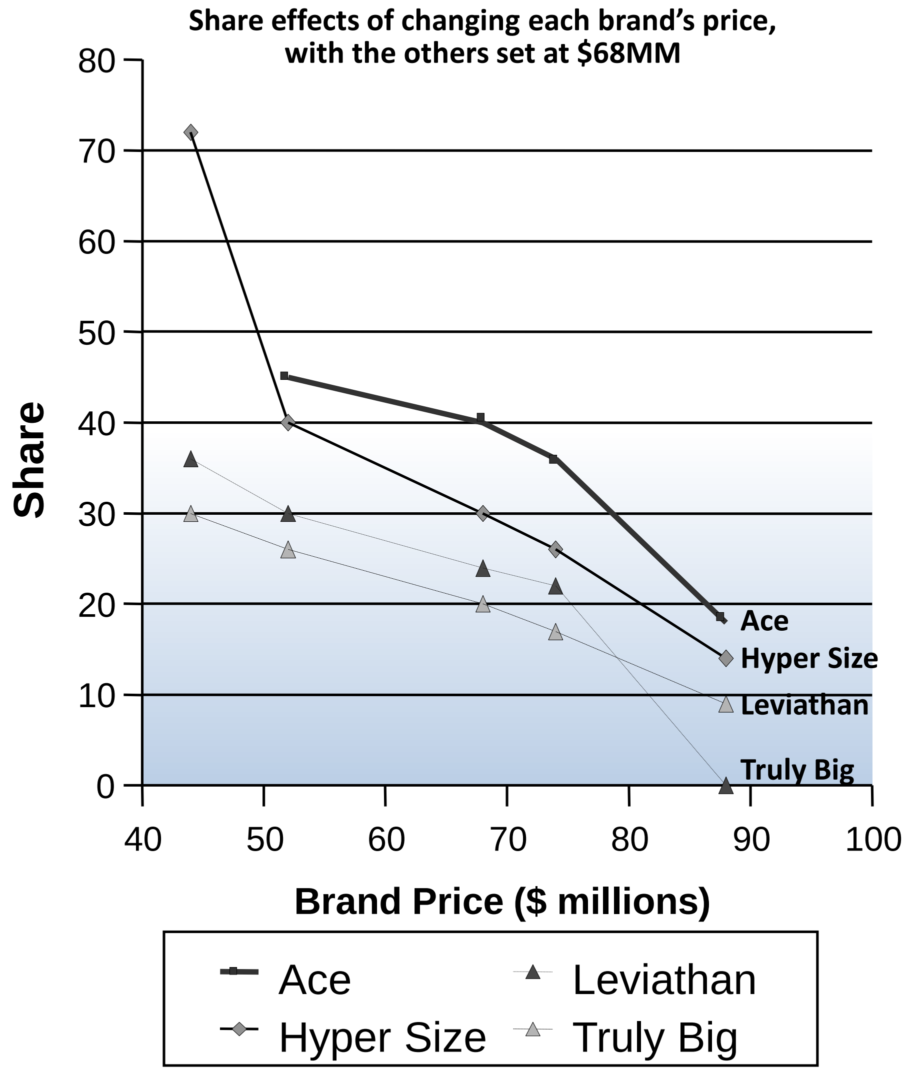
Base case share and share effects of varying one feature at a time



Base case share and share effects of changing attributes

Some helpful DCM charts: The self-effects chart

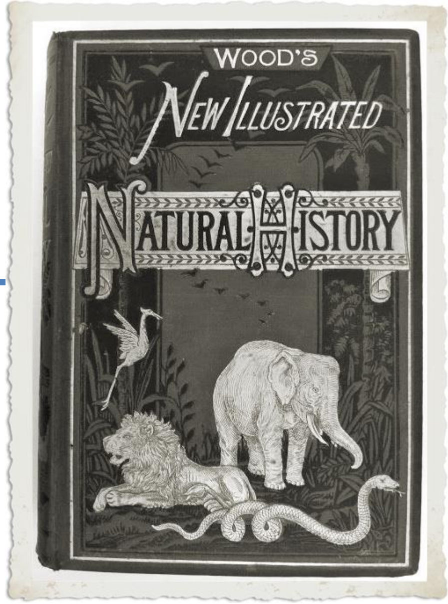
- This chart shows what would happen if each brand varied its price while all others remained at their base level
  - e.g., for all = $68 million
  - For Ace, we see how share would change if all other brands stayed at $68 million and Ace alone changed prices
    - Note that Ace alone does not go below $52 million in price
      - This is below the range Ace's management would consider
  - Superimposing curves for all the brands shows their relative sensitivity to changes in price
    - Note that this one chart summarizes the results of 19 simulations, including the base cases



Converge Analytic

How the methods developed

A very short history



Converge Analytic

Thurstone began all of this, with Case 5

- Thurstone's work was done in the 1920s
- He developed a method to turn rankings of different items into ratio-level scaled data
  - Because this was still early days, he got to call his scaling procedure **the law of comparative judgments** (nothing is the law any more)
  - The procedure for solving it was called **Case 5**
  - This follows guided sorting—it works only at the group level
    - You need to find how many times each item ranks higher than each other
    - This produces a so-called **win-loss** matrix

Part of a "win/loss" matrix. For instance, A wins against B 60 and B wins against A 69 times: Our second most boring illustration

	A	B	C	D	E	F	G
A		69	75	65	79	78	72
B	60		69	59	71	73	68
C	53	60		55	74	67	60
D	63	69	74		83	74	71
E	49	58	54	46		58	52
F	51	56	61	55	71		60
G	57	60	69	58	77	68	

Toward MaxDiff: Pairwise trade-offs work very much like rankings

- There is some confusion about what constitutes **MaxDiff** scaling
- We are taking the common research usage, in which people choose the best and worse—or **best/worst scaling**
- The simplest form is asking which of two items a person likes better
  - i.e., pairwise tradeoffs
- These developed at the same time as Thurstone's rankings
  - They also produce a win/loss matrix
- For many years, like the rankings, these gave only group-level data
  - Commercial **MaxDiff software** developed an extension of this method, allowing 3 to 6 items at a time to get compared
  - Thanks to the near-magic of Hierarchical Bayesian (HB) analysis, we now can get individual-level data from these analyses

There is no other reasonable picture about a win/loss matrix, so here is something amazing by Tintoretto

Converge Analytic

Conjoint developed in market research

- Conjoint was developed in the 1970s by market researchers, largely due to frustration with the poor predictive ability of scaled ratings
  - Scaled ratings **really** do not work well in nearly all instances
  - A possible exception—
    - If you have a lot of historical sales data, a lot of historical ratings data, and the category is not changing, and its buyers are not changing
    - As you might guess—not too likely
- Conjoint proved it was better in real world applications, even in its earliest incarnations
  - It was rapidly and widely adopted

Few purchase patterns or preferences remain the same for long—so historical data often does not work

Early conjoint was not like today's

- Early conjoint (around 1970) looked a little like magic squares—
  - People put numbers in boxes ranking pairs of attribute levels as in the sample below

		Horsepower		
		120	150	180
MPG	30	9	8	7
	40	6	3	4
	50	5	2	1

Our guest respondent filled this out:
1 is best and 9 is worst
Maybe ***this*** *is the most boring illustration?*

- They would do another grid like this for horsepower vs. time 0 to 60, then another for MPG vs. time 0 to 60, etc.

- This still was pretty distant from what people do when selecting a product or service, so development of this method continued . . .

The big development: Full profile conjoint analysis

- Full-profile conjoint arrived in the mid-1970s
- It shows a series of **whole products or services**
  - Hence the name **full profile**
- Respondents rate these product profiles, or (very rarely now) sort and rank them
- This was immediately hailed as a great advance and gained widespread adoption
- It seemed to work well with widely known brands that were similar to each other
- But it also broke down mysteriously in other situations . . .
- And with standard analytical tools, its ability to measure was quite limited

Not our kind of profile, unfortunately

Discrete choice modeling: Econometrics that won a big prize

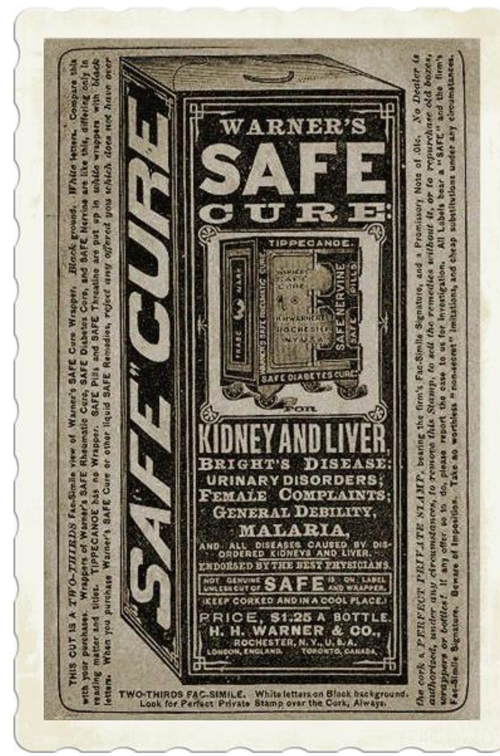
- Work on discrete choice modeling started in the 1960s
  - McFadden eventually won a **Nobel Prize** in economics for this work
- The first widely cited application of discrete choice modeling, published around 1980, answered this question—
  - How we can predict choices when the alternatives do not have any attributes in common
    - This was in transportation, where the choices were taking a train, bus or car to work
    - No common attributes except time door-to-door
  - This study worked!
- And indeed, this method has a remarkably strong track record as well as excellent theoretical underpinnings

A Nobel prize
Nothing else that we are discussing can lay claim to one of these

Converge Analytic

Developments and mutations abound for choice and conjoint

- Many modifications of conjoint have been proposed
  - Best known are partial profile and adaptive conjoint (ACA)
    - Designed to make tasks smaller, reduce attributes measured
    - ACA was roundly criticized and deserved it
    - Some luck here: their popularity may be waning due to increasing use of discrete choice and Hierarchal Bayesian analysis
- Choice modeling has had many proposed extensions as well
  - You may see the commercial products, menu-based choice and (yes) adaptive choice modeling
  - Reports on those methods are mixed
  - Caveat emptor



We suppose this means that this is SAFE

Converge Analytic

More in-depth: Comparing the trade-off methods

Converge Analytic

What is Thurstone’s Case 5?

- In some circles this is very well-known
  - Applied in developmental psychology since 1930
- Published reports show this working with **100 attributes**
  - We have successfully tried 88 (exclamation point!)
- Results look very much like MaxDiff, only with no individual level importances
- Thurstone was influential in psychometrics for many years
  - His work influenced all trade-off methods

The Thurstones (somewhere in there) and friends having a good time

Exactly what is MaxDiff again?

- MaxDiff is both the name of a piece of software and a name for an established statistical procedure that is different
  - The same company that brought us CBC, which conflates elements of conjoint and choice modeling, caused this confusion
- MaxDiff software does this—
  - It starts with a list of items
  - It generates a special experimental design
    - This design makes sure that items are compared with each other in a balanced way
    - It takes the data gathered from respondents and the design and prepares a file that can be analyzed by special HB (Hierarchical Bayesian analysis) software
  - The HB software then generates individual-level data

Sometimes this gets frustrating

Comparing methods: Q-Sort strengths

- Strengths:
  - Study participants rank items, which forces choices
  - Works very well for selecting the favorite item from a very long list
    - e.g., “Which of these 88 fine bonus items would you like best for signing up with our cable service?”
  - Also works very well developing rankings for non-product-related importances e.g.:
    - Service attributes
    - Corporate attributes
    - Communications points, etc.
  - Easy to administer!
  - Easy on respondents
  - Simple analysis

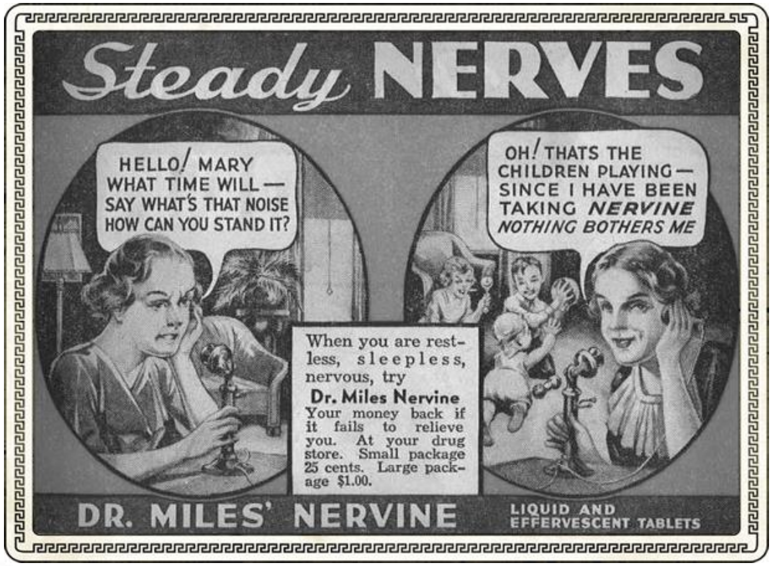
This communication had absolutely no points to prioritize

Converge Analytic

Comparing methods: MaxDiff strengths

- Strengths:
  - Study participants must choose
  - Works very well for selecting the favorite item from a fairly long list
    - Still works well with about 25 items, can be pushed to about 32
    - e.g., “Which of these claims about our custom floor-standing wine cooler are most important to you?”
  - Provides individual-level data
  - Also works very well developing rankings for non-product-related importances e.g.:
    - Communication points
    - Service attributes
    - Corporate attributes, etc.

When promoting your fine product, as in this ad for cocaine, you need to know the important points to get across



MaxDiff and Q-Sort: cons

- Areas of weakness and pitfalls to avoid
  - All items compared must be positive in nature, or all negative
    - You will get tripped up entirely if mixing good and bad
  - Cannot show an entire product but rather separate attributes
  - Cannot address the dynamics of competing in a marketplace
  - Cannot include "must have" items in comparisons
    - They wipe out all other attributes
    - Do not put, e.g., "secure" into a comparison with other attributes for online banking**
  - And remember: you cannot compare **levels of the same attribute**
  - Another **Q-sort con**: no individual level-data

It pays to know what to watch for

***A client did this and "secure" soaked up 99.993% of all importance*

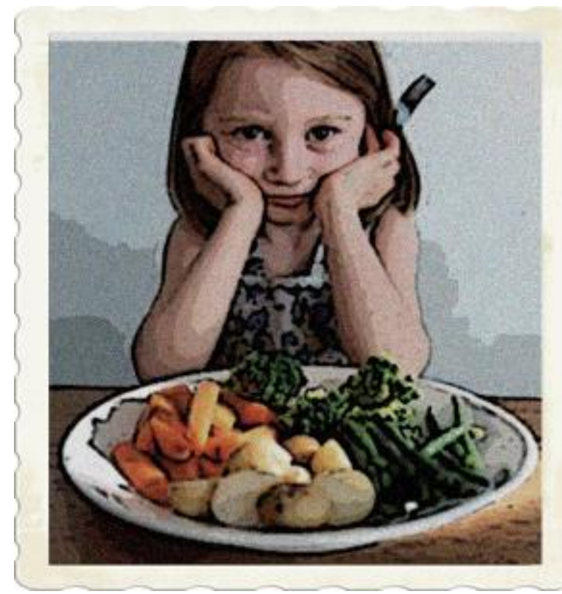
Comparing methods: Strengths of full-profile conjoint

- Relative impacts of different features (and prices) within a whole product are **isolated and measured**
  - That, is measured with no contamination from effects of other features
- Measurement is on a person-by-person basis
- Respondents **must trade off** benefits, just as in the real world
- Since products typically are presented as complete profiles, conjoint greatly increases the **realism** of evaluations
- **Price must be traded** against other valuable features, reducing over-sensitivity of responses to pricing
  - However, some recent evidence shows the effects of price may be **underestimated** by conjoint
    - This may be related to the way conjoint tasks are set up and presented
      - Much more to see/consider in non-price features

Sometimes isolation is better

Comparing: Full-profile conjoint cons

- Conjoint setup demands that all features and prices must appear in connection with all products (see pp. 77-78 for particulars)
  - With branded products, this can lead to unrealistic combinations
    - e.g., a $9,000 Mercedes and $50,000 Yugo in a car study
    - e.g., the antihistamine that puts you to sleep, but with which it is safe to drive a car
- Does not allow you to model the effects of a truly proprietary feature (one that competitors cannot duplicate)
- Not entirely related to marketplace behavior
  - Conjoint asks for preferences, not choices
  - Attitudes often may not equal actions
    - e.g., The "I love spinach" problem**



*** For the benefit of those not recently having an 8-year-old at home to tell this joke: A little girl is invited to a friend's house for dinner. The friend's mother, anxious about wasted food, calls and asks the girl if she likes spinach. The little girl answers, "I love spinach!" When dinner is done, the little girl's plate is still full of spinach. The friend's mother asks what is wrong, since the girl told her she loved spinach. "I do love spinach," says the little girl. "I just don't love it enough to eat it."*

Converge Analytic

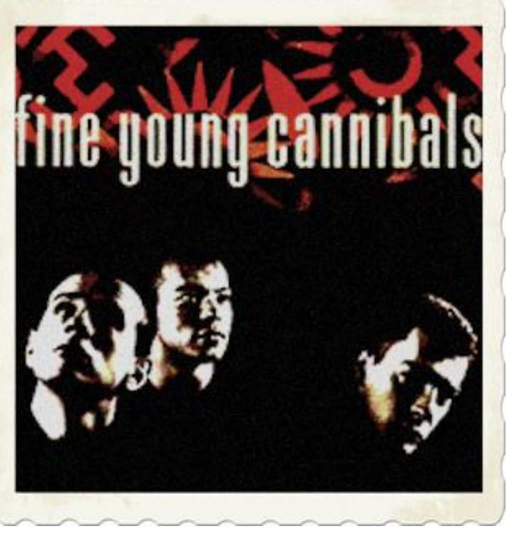
Conjoint disadvantages—models can miss important details

- Conjoint makes some assumptions that are unrealistically simple
  - Critical: Assuming features have the same value in the context of all brands
    - The largest problem: Assuming all brands have the same sensitivity to price changes
    - Alternatively, that all brands have the same **price elasticity**
    - Because brand typically **interacts** with price—some brands can be priced higher—results usually are not accurate for any brand

A general impression may not cut it

Discrete choice modeling (DCM): Strength in more realism

- First and foremost, greater realism—
  - Asking respondents to choose, not give ratings/rankings, and
  - Providing a competitive context for the choices made
- Products can have their own features and prices
  - All features do not have to appear with all products
- More than one product from a brand can appear, modeling within-brand effects—
  - Cannibalization
  - Product line synergies, etc.
- A single model leads both to utilities and share estimates
  - No need for assumptions about how utilities become shares, a problem with conjoint analysis



Not really our type of cannibals

Converge Analytic

Choice-based modeling: Disadvantage in complexity

- Greater complexity
  - We must think of our products and competitors at the same time
  - Could require some understanding of how competitors might change
- Very small brands may get lost in the shuffle
- Has stricter rules than conjoint
  - Conjoint, being looser, ***might*** allow more for unplanned analyses
    - However, you also can make inconsistent or illogical conjoint models—so care is still needed

Consistency really is important

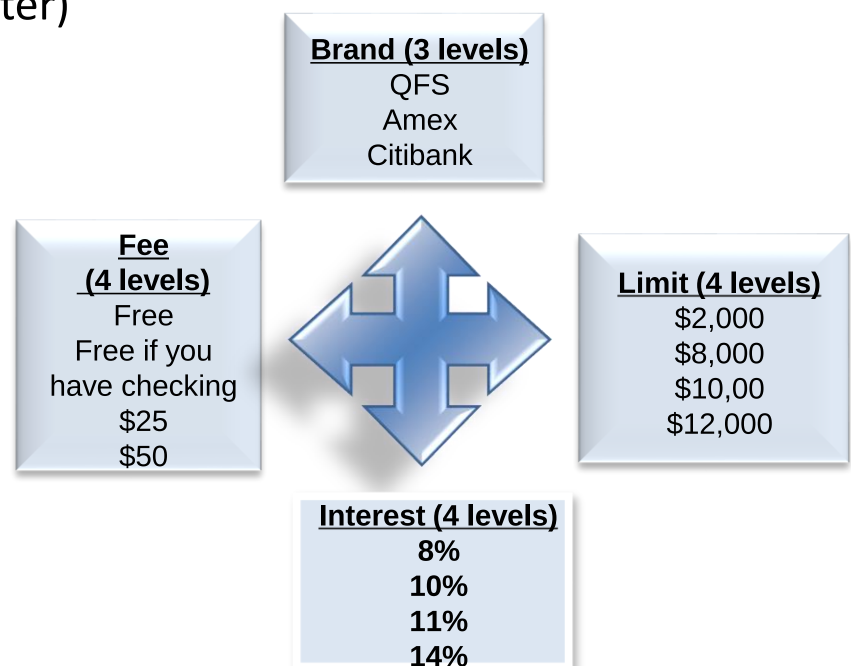
Conjoint vs. discrete choice modeling

More detail on similarities and differences

Which to use when

Choice-based modeling vs. conjoint: Differences in handling brands

- In conjoint, brand is considered an attribute, just like other attributes
  - These all come together to form the value of a product, as the exceedingly clever diagram below shows
- Some attributes may not apply to all brands, but appear with those brands anyhow
  - You may not be able to estimate real market conditions unless you use a large, special design that allows you to investigate interactions (which we discuss later)



Choice modeling: Traditional DCM thinking

- In choice modeling, brand does not even need to become a variable
- Instead, it becomes a **container** holding other variables
  - Variables can be entirely specific to one brand
  - Alternatively, they can apply to several brands, or even to all brands
- Brand comes along for free if we remember that constant term we mentioned way back on slide 42

Variables are specific to each choice

	QFS	Amex	Citibank
Interest	QFS Interest 10%,11%, 13%	Amex Interest 12%,14%	Citibank Interest 8%,11%,12 %
Fee	QFS Fee \$25, \$40	Amex Fee \$35, \$50	Constant
Limit	QFS Limit \$2000, \$8000	Amex Limit \$5000, \$10000	Citibank Limit \$9000, \$12000

Summary: Conjoint thinking vs. traditional choice-based thinking

Conjoint

- Brand is an attribute
- Other attributes apply across all brands
  - They may or may not fit well

Brand (3 levels)
QFS
Amex
Citibank

Fee (4 levels)
Free
Free if you have checking
$25
$50

Limit (4 levels)
$2,000
$8,000
$10,00
$12,000

Interest (4 levels)
8%
10%
11%
14%

Choice-based modeling

- Brand is a container holding attributes
- Attributes are specific to each choice

	QFS	Amex	Citibank
Interest	QFS Interest 10%,11%, 13%	Amex Interest 12%,14%	Citibank Interest 8%,11%,12%
Fee	QFS Fee $25, $40	Amex Fee $35, $50	Constant
Limit	QFS Limit $2000, $8000	Amex Limit $5000, $10000	Citibank Limit $9000, $12000

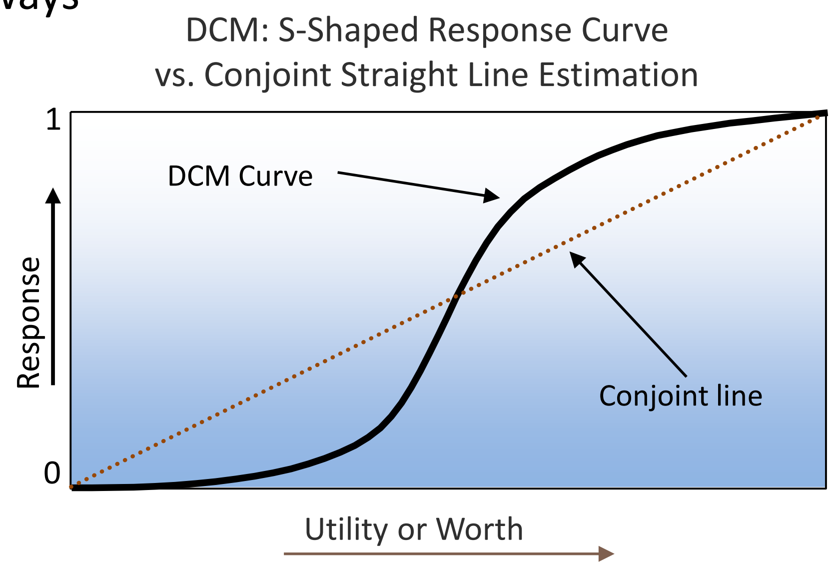
Understanding a little about utility and share

- Both DCM and conjoint produce **utility values** for each level of each attribute studied
- **Utility** is an abstract measurement that tries to capture the exact value of a level of a feature
  - Utility eventually becomes **share of preference**
    - Share of preference is the share a choice gets **in the study**
    - **Share of preference is not market share**
    - To get to **market share** you must factor in—
      - Awareness of the product
      - Understanding of the product
      - Distribution of the product
    - When you adjust for all these, and if you did your study well, you will predict market shares accurately

Definitely not our type of utility

DCM is more realistic than conjoint about utility

- Basic to conjoint: utilities are linear and additive
  - More utility = more preference, in a straight-line relationship
- DCM assumes an **S-shaped** response curve
  - This is more realistic but makes calculations more difficult—
    - You cannot know the value of an alternative just by summing its utilities
    - The curve is nearly linear over the middle range of utilities, though
    - Shares over (about) 60% and under (about) 10% can act in strongly non-linear ways

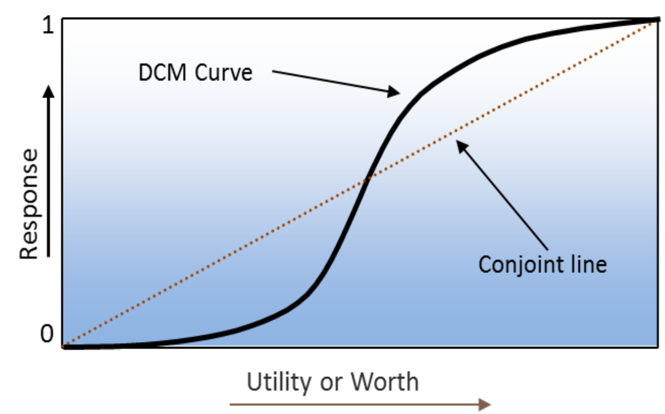


Note: 0 to 1 on the vertical axis corresponds to 0% up to 100%

Now why does DCM take a more realistic view?

- DCM captures marketplace behavior more accurately
  - Utility must pass a certain threshold to get a noticeable response
  - Then small increments in utility boost response strongly
  - Finally, saturation is reached—big boosts in utility are needed to approach a complete (or unanimous) response level
- DCM also reflects the way people respond to stimuli
  - Think of a light slowly getting brighter
    - Up to a certain point, you cannot see any differences
    - After a certain threshold, small differences register strongly
    - When the light becomes too bright, increases in intensity will not register

DCM: S-Shaped Response Curve
vs. Conjoint Straight Line Estimation



This also looks a lot like the decision function in Prospect Theory, one of the latest, newest things.
We will spare you that theory, thankfully.

Discrete choice modeling vs. standard conjoint: Similar basics in goals

- Similarities
  - In the larger scheme, choice-based modeling and conjoint are more similar than different because both are—
    - Multi-attribute trade-off techniques
    - Based on using designed experiments
  - In both, respondents cannot say everything is important
    - They must make realistic decisions, as in the real world
  - Both (usually) attempt to look at complete products
    - DCM, though, also puts products into a **competitive context**
  - Both lead to market share estimates, and can simulate many marketplace situations not explicitly tested
  - Both have plenty of papers backing their use and saying they are generally great stuff
    - DCM has more real-world consistency and even more theoretical justification than conjoint, though
      - Just a reminder that conjoint is still waiting for its Nobel prize
        - Do not hold your breath

McFadden got one of these for discrete choice modeling.

Conjoint versus choice-based modeling: Which is best?

- There have been many arguments, and probably still are some, about whether conjoint or choice modeling is “better”
- Experience shows that each approach works better with certain questions

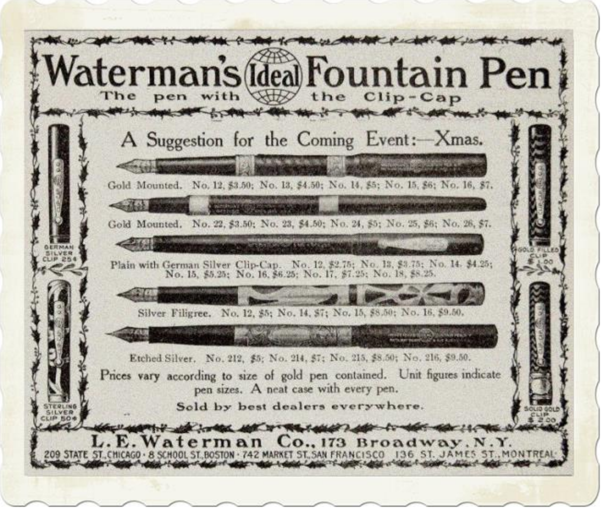
The only saying on a placard you will see here

Let’s not argue about which "works best," but rather understand that each does some things very well, and understand when each works better

- Following, some suggestions . . .

Use full-profile conjoint to optimize single products

- Consider standard conjoint when competitive context is not important
- For instance, you need **to optimize features within a product** (not considering the effects of competition)
  - For instance, this **18 pen** example, optimizing the feel of a disposable pen
    - The goal: get the **best possible writing experience**, without worrying about the competition
    - Pens had 5 features that each could vary 3 ways and one that could vary 6 ways
      - e.g., barrel width, roller ball composition, ink viscosity, etc.
      - There could be some 1458 different configurations
      - An experimental design led to18 prototype pens to be used in testing
      - People tried and rated these
      - Output was a simulator accurately showing the relative appeal of all 1458 possible pens



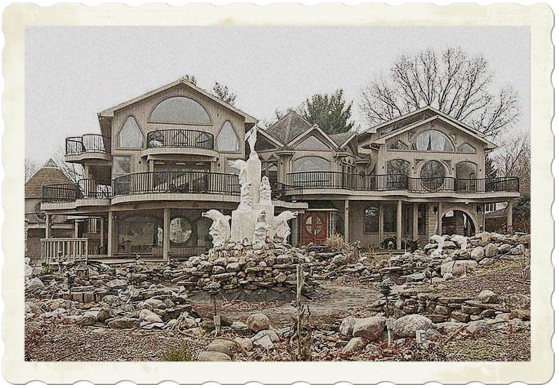
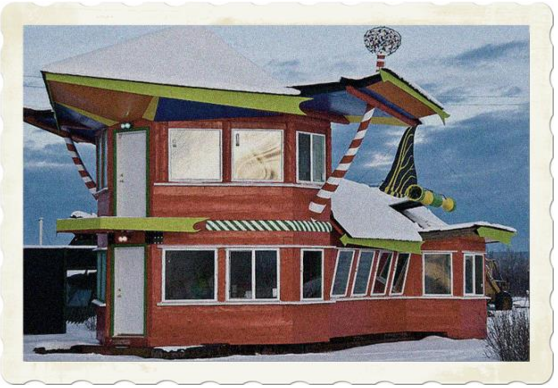
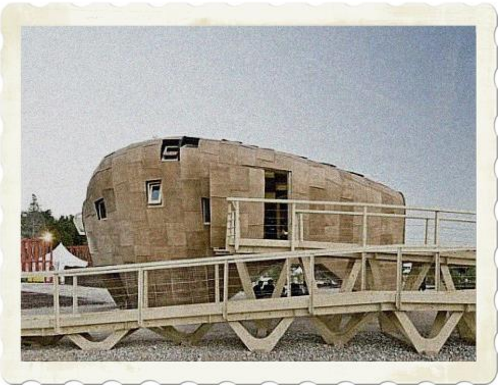
If only they could have done this test back then
(And $9.50 for a pen in the 1920s!)

Standard conjoint for products with no competitors or small shares

- Consider conjoint when **there is no true competitive context for your product**
  - Sometimes products do not have true competitors—for instance, direct mail insurance offerings
    - People indeed do not keep folders filled with old offers to compare
- Consider conjoint when **your product will get chosen only infrequently or owns the market**
  - DCM may not capture the dynamics of highly infrequent choices well
    - If your brand (or others) hardly ever get chosen, then choice models may not pick up what drives these choices
- If your brand is rarely chosen from a wide field, consider narrowing the scope of competition and get a better answer
  - Example: Suppose your product is a breakfast cereal called SoggyOs
    - You could have a 0.5% share of the market and still make a lot of money
    - If you need to model a move vs. competitors, you might look only at specific brands which are your close competitors
      - e.g., other fine shredded cellulose food-like substances
    - Knowing the market really can help

Choice modeling for realism, for allowing “I don’t want any”

- Consider choice-based modeling when—
  - You need the **realism** of evaluating products in their competitive context
  - You expect **different responses to features and prices for different brands**
  - Including both brands and features (or prices) leads to impossible combinations
    - All two-way combinations must appear in standard conjoint
      - However, some methods allows you to include some “prohibitions”
      - Too many of these, though, weaken or undermine results
  - **You want or need to include a "none of these"** choice option
    - This matters in a lot of marketplaces—all choices can be unacceptable

“I think I will choose ‘None of these houses’”

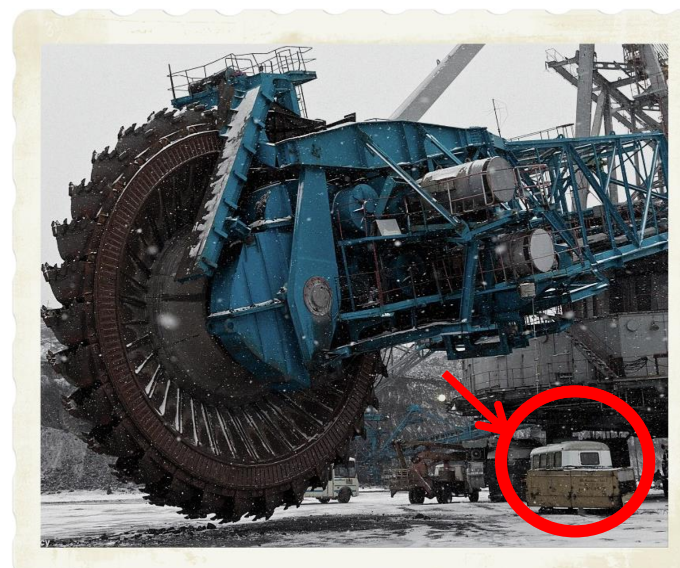
Converge Analytic

Choice modeling for realism: Context usually matters a lot

- Consider choice-based modeling when—
  - You need the realism of evaluating products in their competitive context
    - **Context is critical** for people who are not experts
    - People use the information shown about competitors to anchor their choices
      - Answers without context may grossly mismatch marketplace behavior

Does context matter?
We know the digging wheel to the left must be large—but HOW large only becomes apparent when we have clues from its context.

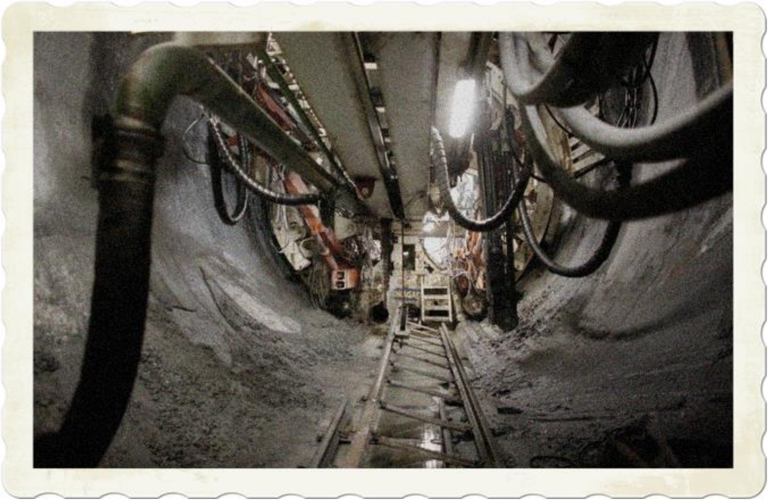
In the photo to the right, that small item we highlighted is some sort of a van or bus. The machine really is that gigantic



DCM when you must have features specific to choices

- Only DCM will be accurate when you have attributes that must be specific to the different brands or choices
  - The way DCM models attributes as existing within each choice is the key
- Similarly, you need DCM when all features or prices tested together lead to impossible combinations for some brands or choices
  - Various somewhat questionable ways to get around this have been proposed for conjoint-like approaches, but some of us do not want to go to those places

Some places, even angels fear to tread



Converge Analytic

And what about this and the other variants?

- This is a screen from CBC (choice based conjoint)
  - It is aptly named, as it works like traditional conjoint, but asks for a choice

Which of the following cell phone plans do you prefer?

1400 minutes	1000 minutes	800 minutes
No rollover of unused minutes	Unused minutes rollover for one year	Unused minutes rollover for one month
New phone every year	New phone every two years	No new phone
Free calling to top 5 contacts	Free nights and weekends (do not count toward monthly minutes)	Nights and weekends count toward monthly minutes
Costs $100 per month	Costs $60 per month	Costs $75 per month
○	○	○

  - Note that brand is missing
  - Not as visible: All the choices have the same attributes
    - No chance to see if an attribute level works better within a specific branded option
    - No chance to test any feature unique to a given carrier
      - e.g., "Most reliable cellular network" (There can only be one of those)

What is not good and reasonable with that choice we just saw?

Not good

- It is not a realistic representation of a marketplace decision
  - We never encounter three unbranded choices
- The context for evaluating prices and features also is likely to be inaccurate
  - Does this range cover the whole marketplace, and if so, are all those prices and features feasible for the client's brand?
    - Presumably, your brand should test only what you can offer
    - What if there is something in the marketplace the client's brand cannot offer, and it is either included in the client's brand or excluded entirely?

Reasonable

- People are at least choosing, rather than rating

What you get from this

- A hierarchy of feature importances, and relatively desirability of each level in the abstract
- However, no sense of what effects might be in the actual marketplace

What about the other methods?

- A number of variants to DCM have been proposed
  - Most notably—
    - Adaptive choice based modeling (ACBM)
    - Menu-based choice modeling (CBCM)
      - From the same company that brought us CBC (choice-based conjoint)
- These are still unsubstantiated by academic research published in peer-reviewed journals
- Still relatively little experience
  - So the verdict is still out: Snow or not?

Your author tends to trust most what has been well tested

- The methods we explain here all have at least some theoretical support and in-market testing
  - Q-Sort/Thurstone Case 5, conjoint and discrete choice (DCM) have plenty of backing
  - MaxDiff has at least some
  - Discrete choice has the most support—including a Nobel prize

Converge Analytic

One kind of summary that may work for you

- Q Sort/Case 5
  - MaxDiff

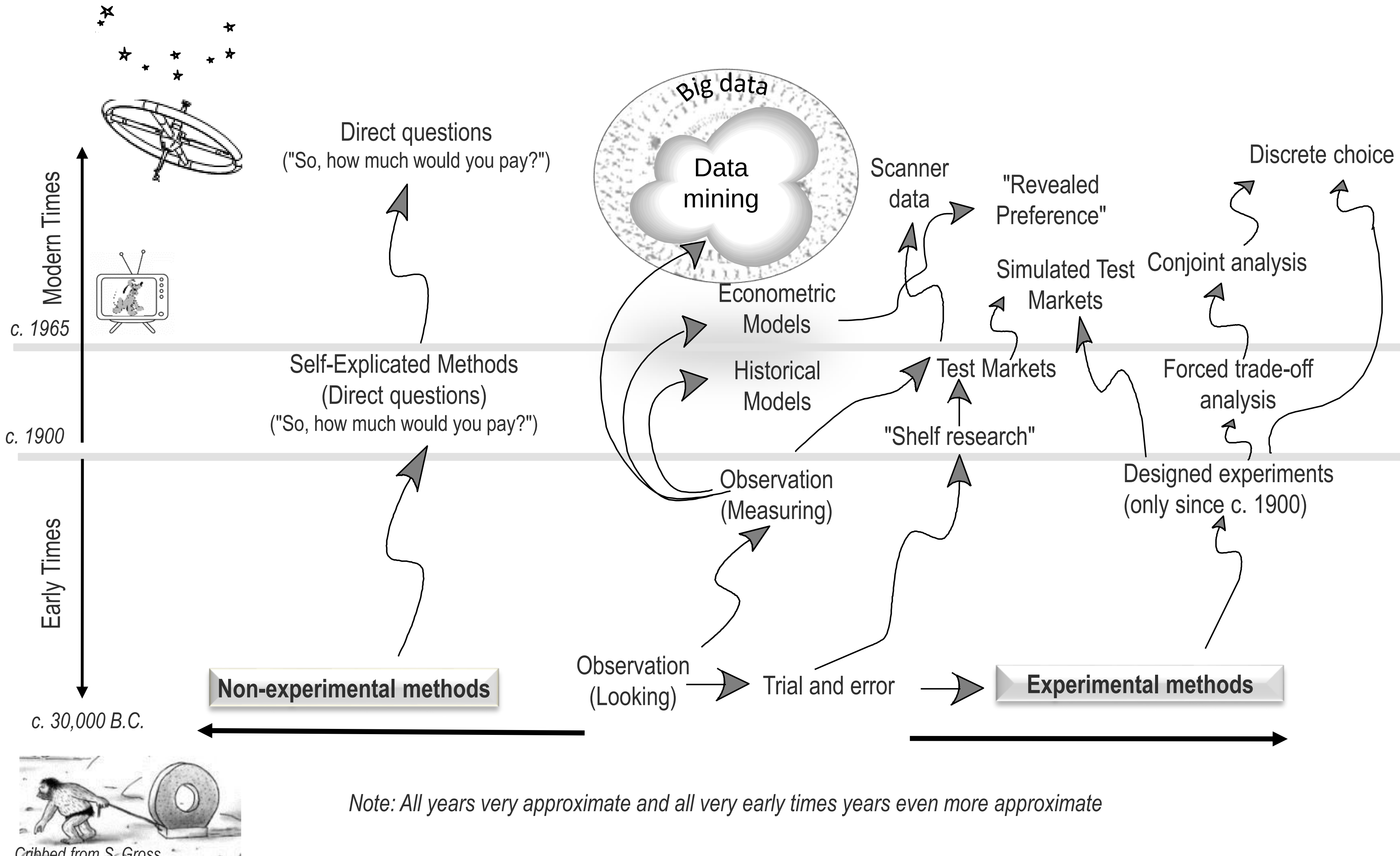
- Full profile conjoint
  - CBC

- Discrete choice modeling (DCM)

Less context	*More context*
More focus on parts rather than the whole	*More holistic*
Less like an entire decision	*More realism*
Easier set up	*More thought required*
Less analytical complexity	*More effort required*
Less effort required	*More analytical complexity*
No market simulations possible	*Accurate market simulations*

A slide intentionally meant to be non-serious

Or choice and conjoint are not so terribly different in the larger context



Note: All years very approximate and all very early times years even more approximate

Appendix

A little further into experimental designs

Side-bar on interactions

Still more methods

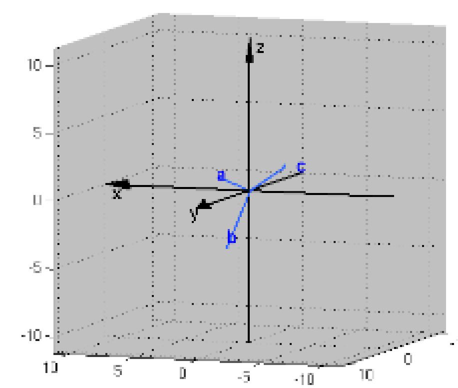
A little further into experimental designs

Basics and suggestions for application

The designs we use: Orthogonal and D-optimal

- Many types of experimental designs have been devised—just a few are highly important for choice models, conjoint and MaxDiff
  - The type traditionally used in choice models and conjoint is called an **orthogonal fractional factorial** design
    - The design discussed on the next page is **fractional factorial**
    - To review—
      - Designs like this let us estimate many the worth or utility of many attribute levels, using very few marketplaces or product descriptions
      - However, they sacrifice something to get this great efficiency
        - They are not designed to measure *interactions* among attributes
      - Interactions will be coming up soon
  - Another design, usually aided by computers, is similar to the fractional factorial type—the **D-optimal design**
    - These may provide real benefits in reducing the size of a design vs. a standard fractional factorial

Not to confuse things unduly, but axes at right angles (like x, y and z here) also are called orthogonal—and they indeed have zero correlation



Converge Analytic

A standard orthogonal fractional factorial design

- That is a mouthful! So much so that one former boss used to charge an extra $500 for a "certified orthogonal fractional factorial design"
- Let's recall the basic rules—
  - Each **attribute** appears as a **variable** (column)
  - Levels are encoded (in this case starting with zero; some programs start with one)
    - e.g., three levels for one attribute would appear in the design as 0, 1, 2 in one column
- Each **card** or **screen** (or product profile) will be one row of the design
  - Reading across the row gives one product's configuration

	Attribute A	Attribute B	Attribute C
Card 1	0	0	0
Card 2	0	0	1
Card 3	0	1	0
Card 4	0	1	1
Card 5	1	0	0
Card 6	1	0	1
Card 7	1	1	0
Card 8	1	1	1

You see how the attributes change from one card (or screen) to the next by reading down the columns

*This setup with the attributes as columns means that the attributes are the **variables** in the design*

- Each attribute level appears at least once with each other attribute level from each other attribute
  - That is, all **pairwise combinations** are covered, but the design may not cover all possible three-way combinations of attribute levels

More about designs: orthogonal means no correlations

- To review, we use correlations among the columns as the standard measure of how closely they are related
- Note that the way in which each attribute is varied (from one card, or marketplace, to the next) has absolutely no correlation with the way in which any of the other attributes vary
- **No correlations** means that the way in which any given attribute varies cannot have an influence on the value we get for another attribute
  - This is all that **orthogonal** means: **no correlations** between the variations in any of the attributes

Correlations

		A	B	C
A	Pearson Correlation	1.000	.000	.000
	Sig. (2-tailed)	.	1.000	1.000
	N	8	8	8
B	Pearson Correlation	.000	1.000	.000
	Sig. (2-tailed)	1.000	.	1.000
	N	8	8	8
C	Pearson Correlation	.000	.000	1.000
	Sig. (2-tailed)	1.000	1.000	.
	N	8	8	8

Sorry: none of this involved!

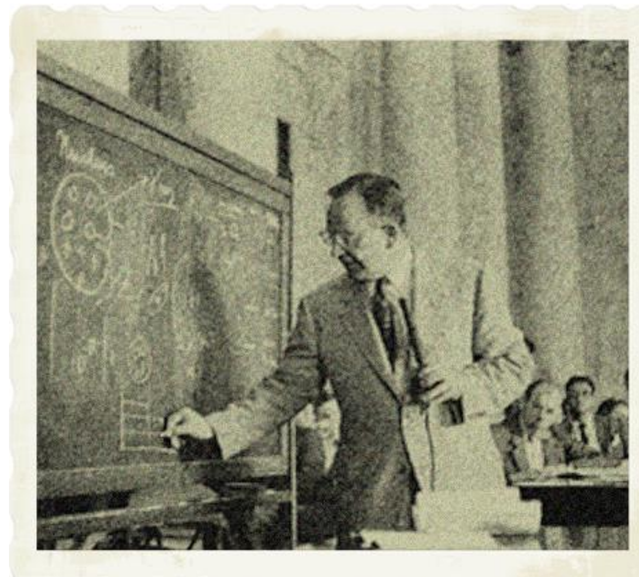
A more difficult fractional factorial design

- Here we have 5 attributes. One has 4 levels (0 to 3), one has 3 levels, and 3 have 2 levels
  - This requires 16 cards or screens (or marketplaces or scenarios)
- This also is a much more difficult balancing act (and not recommended for your spare time)

	Attribute A	Attribute B	Attribute C	Attribute D	Attribute E
Card 1	3	2	1	1	0
Card 2	2	2	0	0	0
Card 3	1	1	0	0	1
Card 4	1	2	1	0	1
Card 5	3	0	0	0	1
Card 6	0	2	0	1	1
Card 7	0	0	1	0	0
Card 8	2	0	0	1	1
Card 9	2	0	1	1	1
Card 10	2	1	1	0	0
Card 11	3	0	1	0	1
Card12	1	0	1	1	0
Card 13	3	1	0	1	0
Card 14	1	0	0	1	0
Card 15	0	0	0	0	0
Card 16	0	1	1	1	1

Once again, all the correlations among the different attributes are zero

		A	B	C	D	E
A	Pearson Correlation	1.000	.000	.000	.000	.000
	Sig. (2-tailed)	.	1.000	1.000	1.000	1.000
	N	16	16	16	16	16
B	Pearson Correlation	.000	1.000	.000	.000	.000
	Sig. (2-tailed)	1.000	.	1.000	1.000	1.000
	N	16	16	16	16	16
C	Pearson Correlation	.000	.000	1.000	.000	.000
	Sig. (2-tailed)	1.000	1.000	.	1.000	1.000
	N	16	16	16	16	16
D	Pearson Correlation	.000	.000	.000	1.000	.000
	Sig. (2-tailed)	1.000	1.000	1.000	.	1.000
	N	16	16	16	16	16
E	Pearson Correlation	.000	.000	.000	.000	1.000
	Sig. (2-tailed)	1.000	1.000	1.000	1.000	.
	N	16	16	16	16	16



*Back when you could have a fun time talking about zero correlations (This is **not** your author)*

Design standards: The full story of how many rows or cards or screens

- We need more screens (or tasks, or marketplaces, or cards) as we measure more attributes and levels
- Here's a quick general check for minimum design size (design saturation, or when the design gets completely full)
  - (Number of attributes X number of levels) - number of attributes + 2
  - **Example:** 8 attributes, three with four levels, 5 with 2 levels:
    - (3 x 4) + (5 x 2) = 22
    - subtract 8 = 14
    - add 1 for error and 1 for measuring the constant term = 16
  - Therefore, select the smallest design that requires 16 screens or cards
- **Note**
  - Some say you need to add 3 instead of 2,

    or multiply the subtracted total (13, in this case)

    by 1.1 and round up
    - However, adding two works fine

Poor designs can hurt you

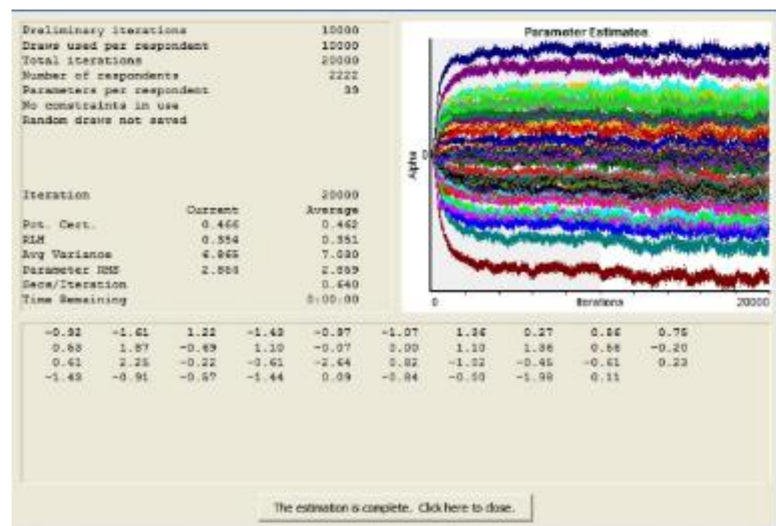
Getting more from data—the world of HB analysis

- We mentioned that HB (Hierarchical Bayesian) can squeeze much more out of a given standard designs
  - You can get **3 or 4 times more information** reliably—amazing but true
  - However, we still need to know 3 or 4 times **more than what** amount
- Fixed design programs usually use catalogues of designs rather than designing from the start
- In these, no design may exist approaching the theoretically lowest size
  - Therefore, if you used fixed designs, the best design you can find may be much larger than **saturation size** (or the design being "full")
  - **D-optimal** designs may help out here
  - However, even D-optimal designs cannot exceed **saturation**
- Thanks to HB, we now can measure as much as a respondent can stand in a study

All the latest HB hits playing here

But what is HB analysis?

- Briefly, it fills in data that is scant or missing for a respondent by repeatedly **borrowing** estimates from other respondents
  - That is, it keeps sampling other respondents and storing values from those who have the missing information
  - It usually runs 20,000 or more times for each attribute level for each respondent, keeping a running average of its estimates
    - It may or may not compare the respondent to the sample it is drawing and make adjustments based on their similarities
  - Estimates will settle down to steady values (or **converge**) if you have set up the problem correctly
    - If you have not, then maybe not
    - A solution that does not converge usually means errors in setup, data collection or coding of values
  - It gives your PC more of a workout than almost anything else
  - You will wait for a complicated DCM run to finish, maybe hours
- Amazingly, **all this borrowing works**—and we get very accurate estimates

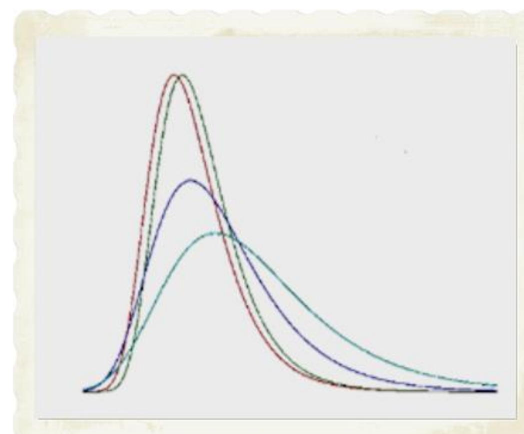


How values vary from 0 to 20,000 estimates (and looking mostly stable at the end)

Review: From design size to sample size

- You need enough sample for the analysis to provide reliable results
- As mentioned, errors for DCM tend to be tighter than the errors around sample percentages—meaning we may get away with a slightly smaller sample for a given level of precision
- For **standard analysis** (no HB added), getting to a minimum workable sample is fairly simple
  - Start by thinking of **125 (full/complete) respondents** as the minimum
    - Both for conjoint analysis and for DCM where people choose one item
    - This means 125 respondents per group you want to analyze separately
  - Next, look at the number of attributes and levels in your design
    - Use your favorite formula to determine how many scenarios/cards that will require
  - Then work up the sample as follows . . .

Here is another shot of the good old Gumbel distribution. Remember this from page 46? This is narrower than our even older friend the "normal" distribution. Narrower error means more precise measurements and so less sample is required



An example getting all the way to sample size

- Say you have 11 attributes with 3 levels and one with 2 levels
- You need (11 * 3 [or 33] +2) – 12, (and add back in 1)--or 24 cards or screens
- This gives you one **replication** or the equivalent of one full respondent
  - Now, suppose you worry that your respondents will tired easily and so you want to give each person in the survey **8 screens** to evaluate
  - That works out to exactly one-third of a full respondent apiece
    - This means you will **need to interview 375** people to get the equivalent of **125** full respondents, with no HB analysis
    - If respondents **allocate** (e.g., over the next 10 patients) and they are a fairly homogeneous population, **you likely can get by with only 75 per group**
- Remember, **with HB analysis** and a good size experiment, you can **get 3 to 4** times more out of each respondent
  - Or, here you likely could get by using **only 125—** getting practically all the power you would with the 375 sample
  - **That is a vast improvement**!

On the Web illustrating "vast improvement" and nicely framed for you too. One never knows

What are random designs?

- A software company (Sawtooth software)[1] has proposed **random designs**[2] for discrete choice (or CBC)
- A computer mixes up the attributes and levels and gives a random combination to each respondent
- With enough respondents, this should cover every combination of attributes and level
  - Some good analysts say this works well
- There could be some concerns about how well it pans out with HB (Hierarchical Bayesian) analysis, though
  - This method fills in spotty data by "borrowing" repeatedly from samples of other respondents
  - It is not clear what it is borrowing if everybody is doing something different from everybody else
  - You rely on everything coming out in the wash
    - In fact, it might do so—perhaps

Another design billed as random, all ready to stick in your photo album

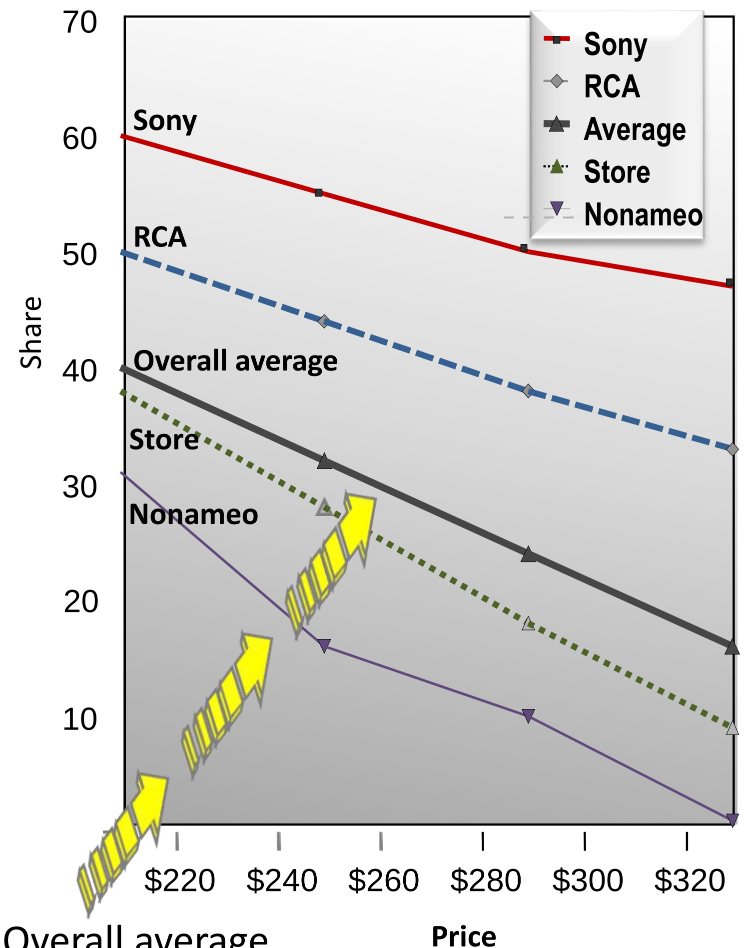
[1]The same folks who brought you CBC, ACBC, etc.

[2]This does seem like an impossible combination of words

Concluding side-bar: What is an interaction?

- **Interaction** has practical and statistical meanings
  - Practically, this means:
    - To get the value of a variable we care about (e.g., market share) we must understand how two or more other variables influence each other
    - That is, we must know how those variables behave together to get accurate readings of share changes
      - Knowing just one is not enough
- **Example** Suppose we have 4 brands of televisions: Sony, RCA, Store Brand, and Nonameo***
  - Suppose each of these brands could be sold at any of 4 prices:
    - $209, $249, $289, and $329
  - If (e.g.) Sony sells better at all prices than the other brands, then brand and price **interact**
  - With conjoint style designs, you must specify that **brand and price interact** to see the different price vs. share response patterns for the 4 brands
    - Otherwise, you get an average (**generic**) price curve that does not fit most of the brands well
- **With a price variable specific to each choice, DCM would eliminate the need for this interaction term**

How share changes in response to changes in price



Overall average
(follow the yellow arrows)

*** *You never heard of it.*

Why we care about interactions: They make for many measurements

- Interactions can blow out the total number of **parameters** or terms you want to measure
  - When added to a model, they greatly increase the number of terms
    - For instance, 3 brand and 4 prices, using an **interaction term adds 12 more terms** to your model
    - You actually need a **total of 19 terms**—3 for brand, 4 for price and 12 for the interaction—this multiplication adds many terms
  - You would do better giving each brand had its own price attributes
    - That would be 12 levels
      - 4 prices x 3 brands
      - But that would be all
    - We do not need to measure brand—it comes along "for free" as a constant in the choice model
    - So **we save 7 terms** in the model
      - And get accurate, direct measurement in the bargain

We have to be careful about multiplying

Converge Analytic

Still more alternatives methods to DCM

A brief look at pros and cons

Trade-offs have many competing methods

- As a reminder, these are the trade-off methods we use now—
  - Q-Sort/Case 5 guided sorting of items
  - MaxDiff forced trade-offs of 2 to 5 items
  - Full profile conjoint analysis (or standard conjoint)
  - Discrete choice modeling (DCM)
  - And some variants of these, such as CBC
- Many, many other proposed alternatives to using trade-offs exist
- We can classify them in many ways, e.g.:
  - Direct questions about what's important
  - Test markets
  - Econometric methods
  - Big data (whatever that is)
  - Miscellaneous methods

We will do our best to sort things out

Self-explicated methods: It's obvious—but answers are inaccurate

- Strengths
  - An obvious-seeming approach
    - Typically just asking somebody how important he/she finds each feature
  - Seems perfectly clear to nearly everybody
    - No mystery in saying, e.g., "90% of respondents said chrome rims are critical"
- Weaknesses
  - Answers often greatly overstate the importance of features
    - It costs nothing to say all features are "extremely" important—or critical
  - Ratings do not differentiate between essentials and the "nice-to-have"

No doubt, with standard rating scales, customers identified each of these as essential

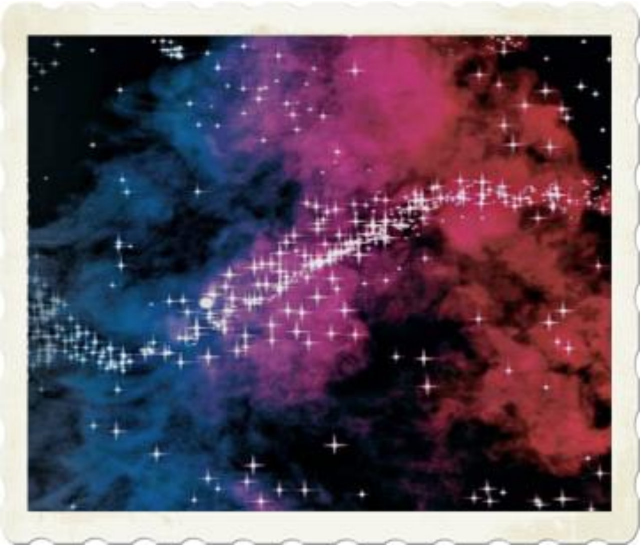
Test markets—Observed but little ability to experiment

- Strengths
  - Full scale test markets definitely provide observational data based on behavior in marketplaces
  - Adoption rates can be directly measured, as can various media-related effects (from alternative advertising, promotions, etc.)
- Weaknesses
  - Logistics and costs of setting up can be staggering
  - Limited ability to manipulate many aspects of the product at the same time
  - Questions about competitive responses which do not occur in the test market cannot be addressed
  - Competitors often discover test markets, and do their best to sabotage them

All you do is make the product, get stores to distribute it, make advertising and promotions, and then advertise and promote, in places like these. Then analyze. Nothing to it!

Econometric methods as alternatives: Pro—Actual data

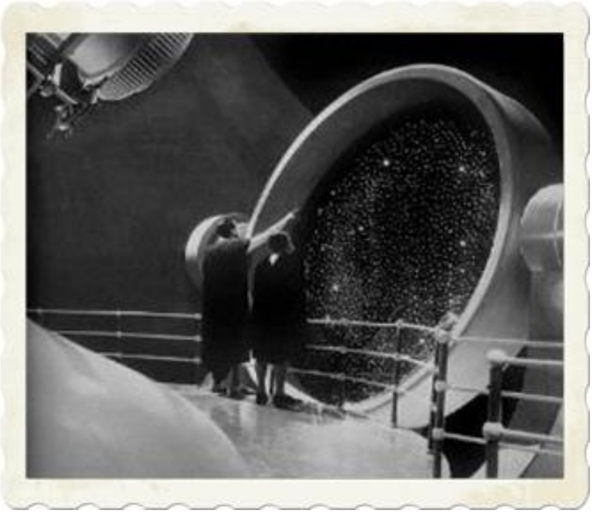
- Involve gathering historical data on sales patterns and trying to find variables that will explain observed changes
- Strengths
  - As historical records of behavior, these definitely reflect what buyers (and non-buyers) did in the marketplace
  - Highly sophisticated modeling techniques have been developed for dealing with historical data



The analytics can be nearly rocket science

Econometric methods: Cons—Limited to history only

- Weaknesses
  - Information on key variables can spotty, inaccurate—or missing
  - Various other motivating factors may be overlooked or forgotten
    - Feedback from actual buyers and non-buyers rarely enters into these models
      - Even when these models show what has happened, they may reveal little about why changes occurred
  - As may be obvious, there is no way to vary or experiment with historical data
  - You are looking backward, not forward



Things to come are not visible in the rear view mirror

And 20 bonus points if you noticed we used this image before (10 more points if you already found it on page 28)

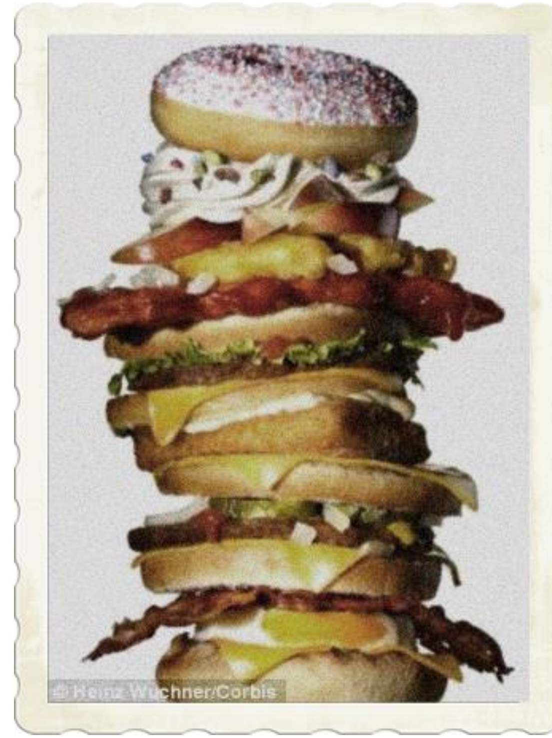
Big data—whatever that is

- Big data is certainly one thing
  - A lot of data
- One definition, only semi-humorous—
  - More data than you can handle with your current equipment
- Big data used to be terabytes of data
  - Now that we can handle those, we may need to move to petabytes
    - Exabytes, Zettabytes and beyond await
- Surprisingly many believe that if you have enough data, you can solve any problem
  - However **data is not information**
    - That should be on a plaque someplace
    - Information helps you deal with a novel situation and make better decisions
    - Data is just stuff

What I want must be in here somewhere

Now what can we expect from big data?

- Pro
  - Real data and lots of it
- Con
  - Answers simply do not emerge from having lots of data
  - Techniques applied to big data can be sloppier and/or more questionable than in econometrics
    - For instance, finding "significant" correlations with massive samples
    - With a large enough sample, everything passes a test of significance
  - Watch reports about big data **very carefully**
    - e.g., Google Flu Trends predicted a flu outbreak correctly in 2008
    - It was wrong 100 out of 108 times from 2011 to 2014
  - Any data not gathered for the purposes you have in mind may not have what you need
    - And like all retrospective data, it has almost no ability to forecast something new



Just because there is a lot of it does not mean it is good for you

Converge Analytic

Miscellaneous methods as alternatives

- Dozens of alternative methods have been proposed as ways to optimize products or pricing
  - These seem to fall into several classes:
    - Mysterious or **black box** approaches
      - Details of how they work are not available or are "proprietary"
      - Many make extravagant-seeming promises
      - With the rise of powerful, carefully reviewed methods, these seem to be waning
    - Unsubstantiated methods
      - These can seem interesting, but no support found anyplace
        - One fairly well-known: van Westendorp PSM
        - After 40+ years, still no published substantiation!
    - Repudiated methods
      - May have had some following but have serious flaws
      - Due to history, they still get some use
        - One fairly well-known: BPTO, or brand price trade off

Finally, please stay away from these and use the powerful methods we have discussed

Questions? Comments? Need more information?

Contact Steven Struhl

smstruhl@convergeanalytic.com

smstruhl@gmail.com

847-624-2268

Thank you for staying until the end—definitely no more slides after this