Experiments for Messaging Research

Keith Chrzan, Sawtooth Software

Sawtooth Software
The survey software of choice

Messaging

- Messages are the words used to convey the value proposition of a
  - Company
  - Brand
  - Product
  - Feature
- Messages include taglines, slogans, ads and product or feature descriptions

Topics

- Messaging unit of interest
  - Individual message elements
  - Complete messages: combinations of message elements
  - Executions
- Questions about messages – how to . . .
  - Prioritize elements/combinations
  - Find a reach-maximizing set of elements
  - Find a utility-maximizing combination of elements
- Approaches for testing messages
  - Monadic tests
  - MaxDiff
  - Best-worst case 2 scaling (best-worst conjoint), with or without a None alternative
  - Choice-based conjoint experiment

TYPES OF MESSAGES

Types of Messaging Research

1. Sometimes we have a large number of message elements that can be combined in various ways to form whole messages – in this case we want to do research to cull the long list of potential message elements down to a short list which can be combined into a message execution
2. Sometimes we want to evaluate a list of complete messages to find a winner
3. Other times we want to test the impact of complete messages on customers' willingness to purchase
4. On occasion we have fully developed message executions (e.g. ads) to test

This presentation focuses on #1 - #3 above – testing of messages or message elements, but not of fully developed ads

Message Elements

- Message elements are the components of a message:
- Product X is easy to use, effective, and won't stain your clothes; coverage lasts 18 hours and Product X saves you money
- Moreover, there may be several ways to communicate that Product X is a good value (i.e. alternative elements may express the category "value")
  - Saves you money
  - Won't break your budget
  - 50% less expensive than the leading competitive product
  - Good value for the money
  - One 12-oz container will last 6 weeks
  - Costs the same today as in your grandpa's day
- Lists of message elements can get long (e.g. 100+ items)

Complete Messages

- Product X is easy to use, is effective and won’t stain your clothes; coverage lasts 18 hours and Product X saves you money.
- Each Hilton Hotel & Resort is a unique reflection of its destination and combines local influences with out-of-this-world service to make your stay truly memorable.
- The 2020 CR-V blends functionality with pure driving bliss. We took a sporty look and packed it with innovative technology and performance for a truly dynamic driving experience.

QUESTIONS ABOUT MESSAGES

Goals – Message Elements

- Do we want to prioritize message elements?
  - Which are the 3 most appealing message elements?
  - What is the most valuable category of message elements?
  - What is the most appealing element in each category?
- Do we want to evaluate how best to combine message elements?
  - Which set has message elements that appeal to the most people?
  - Which combination of message elements will attract the most people?

Goals – Complete Messages

- Do we want to find the most appealing of a longer list of complete messages?
- Do we want to see how a smaller number of complete messages contribute to product preference?

APPROACHES FOR TESTING MESSAGES

MONADIC TEST

Monadic Test

- We could show each respondent one message (or message element) and get detailed ratings about it
  - Overall liking
  - Credibility
  - Appeal
  - Willingness to buy
  - etc.

Benefits of a Monadic Test

- It there are 7 messages we want to test, then we should have 7 cells in our experiment (one per message)
- We recommend at least 200 respondents per cell and preferably 400
- Benefits:
  - You get a good clean read on each message, uncontaminated by the effects of respondents potentially having seen the other messages
  - Analysis is simple - Z-tests of top-two box proportions or perhaps t-tests or ANOVAs on ratings to find significant differences
- Drawback: this method is sample size intensive (for 7 messages, sample size will be at least 1,400 respondents)

Sequential Monadic Test

- Alternatively, we could show each respondent several messages sequentially collecting these ratings in turn for each one
- We control the order of messages with an experimental design
- We recommend at least 400 respondents for this kind of experiment

Benefits of a Sequential Monadic Test

- Benefits:
  - Less demanding than a pure monadic test in terms of sample size
  - You get a good clean read on each message, controlling for the effects of respondents potentially having seen the other messages
  - Analysis is simple - Z-tests of top-two box proportions or perhaps t-tests or ANOVAs on ratings to find significant differences
  - Because of the repeated-measures aspect of the design, your stat tests will be more powerful (e.g. repeated measures ANOVA rather than one-way ANOVA; or dependent t-tests instead of independent t-tests)
- Drawback: this method gets more taxing on the respondent or more complicated for the researcher as the number of messages gets larger

MAXDIFF

MaxDiff

- If we want to focus on prioritizing the messages, we can use a MaxDiff experiment, with several questions like on the next page
- The focus of these questions is very much on the messages themselves and not on any competitive context

Example Question

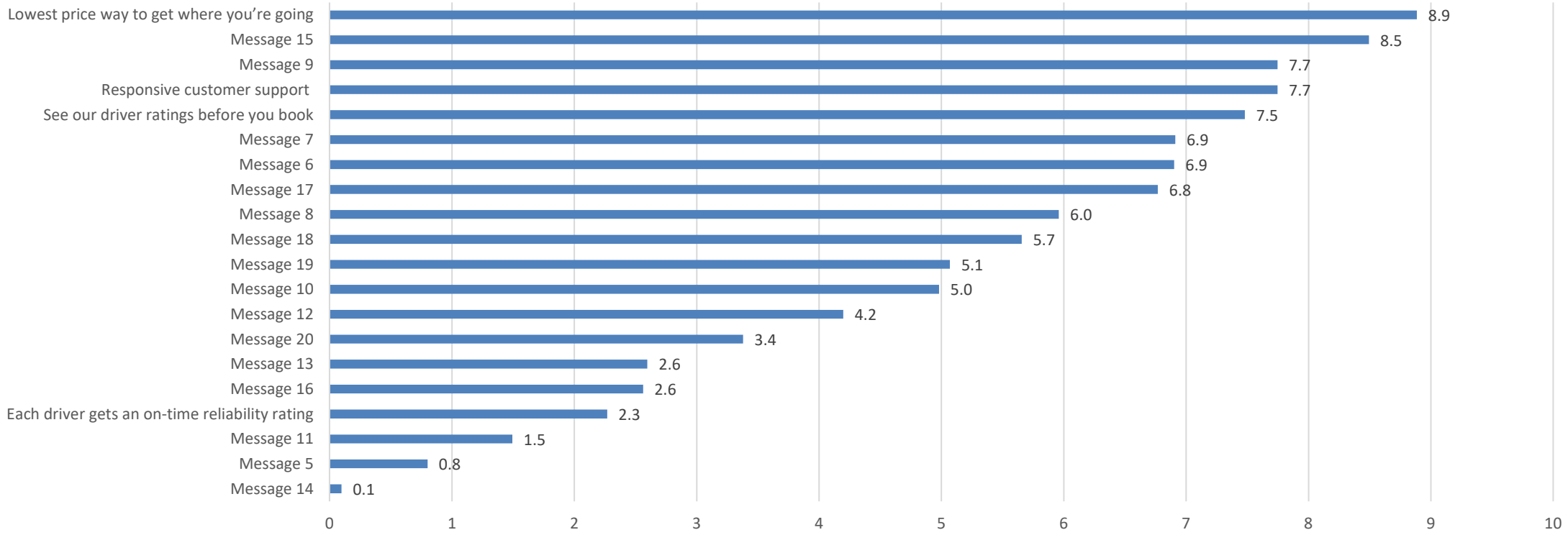
Which one message **most** makes you want to use UltraUber and which message **least** makes you want to use UltraUber?

Most	Message	Least
☐	Lowest price way to get where you're going	☐
☐	See our driver ratings before you book	☐
☐	Responsive customer support	☐
☐	Each driver gets an on-time reliability rating	☐

- If we had to 20 message elements in total, we might ask each respondent 15 questions like this one, each with a different set of 4 messages

Results – Message Prioritization

- We get the utility of each message element (utilities in this case sum to 100)



Prioritization

- Identifies the relative value of each element
  - “Lowest cost way to get where you’re going” is the most valuable message element
  - “Each driver gets an on-time reliability rating” isn’t as valuable as most other elements
- Note that we could do this with entire messages as well as message elements

Results – TURF Analysis

- TURF analysis allows us to find the bundle of (e.g.) 5 messages that together have the best “reach”
- For example, the bundle of messages that gives the most respondents their favorite message is
  - Lowest price way to get where you’re going
  - Message 15
  - Responsive customer support
  - Message 6
  - Message 8
- Why isn’t message 9 (the third ranked message) on this list?

MaxDiff Advantages

- Advantages
  - Survey is simple for respondents
  - Experiment is easy to construct and analyze
  - Utilities sum to 100 and are ratio scaled (so a message with a utility of 18 is twice as valuable as one with a utility of 9)
  - Allows clear prioritization
  - Enables TURF analysis

BEST-WORST CASE 2 SCALING (“BEST-WORST CONJOINT”)

Best-Worst Case 2 Scaling

- Sometimes our message elements fall into categories
- e.g. we might have 25 messages in total
  - 5 message elements about product quality
  - 4 elements concerning ease of use
  - 4 statements about safety
  - 6 covering aspects of price
  - 6 message elements describing durability
- Instead of allowing these to combine freely, we might show MaxDiff questions containing 5 items, one from each category (we control for this in the MaxDiff software by adding prohibitions)

Example Best-Worst Case 2 Question

Which one message **most** makes you want to buy the product and which message **least** makes you want to buy the product?

Most	Message	Least
☐	Quality 1	☐
☐	Ease of Use 2	☐
☐	Safety 4	☐
☐	Price 1	☐
☐	Durability 6	☐

Analysis of Best-Worst Case 2

- We still prioritize elements, but now we can identify, e.g. the top element in each category
- Unlike a conjoint experiment, we can directly compare the utility of levels of different attributes
- We can still run TURF analysis, finding the reach-maximizing bundle of items that are constrained to have one item from each of the 5 categories
- We can also get at the value, or utility, of combinations of items, which we can't do with MaxDiff, by adding a simple follow-up question, as on the next slide

Example Best-Worst Case 2 Question

Which one message **most** makes you want to buy the product and which message **least** makes you want to buy the product?

Most	Message	Least
☐	Quality 1	☐
☐	Ease of Use 2	☐
☐	Safety 4	☐
☐	Price 1	☐
☐	Durability 6	☐

If this product was available where you shop, would you consider buying it or not?

- ❑ Yes
- ❑ No

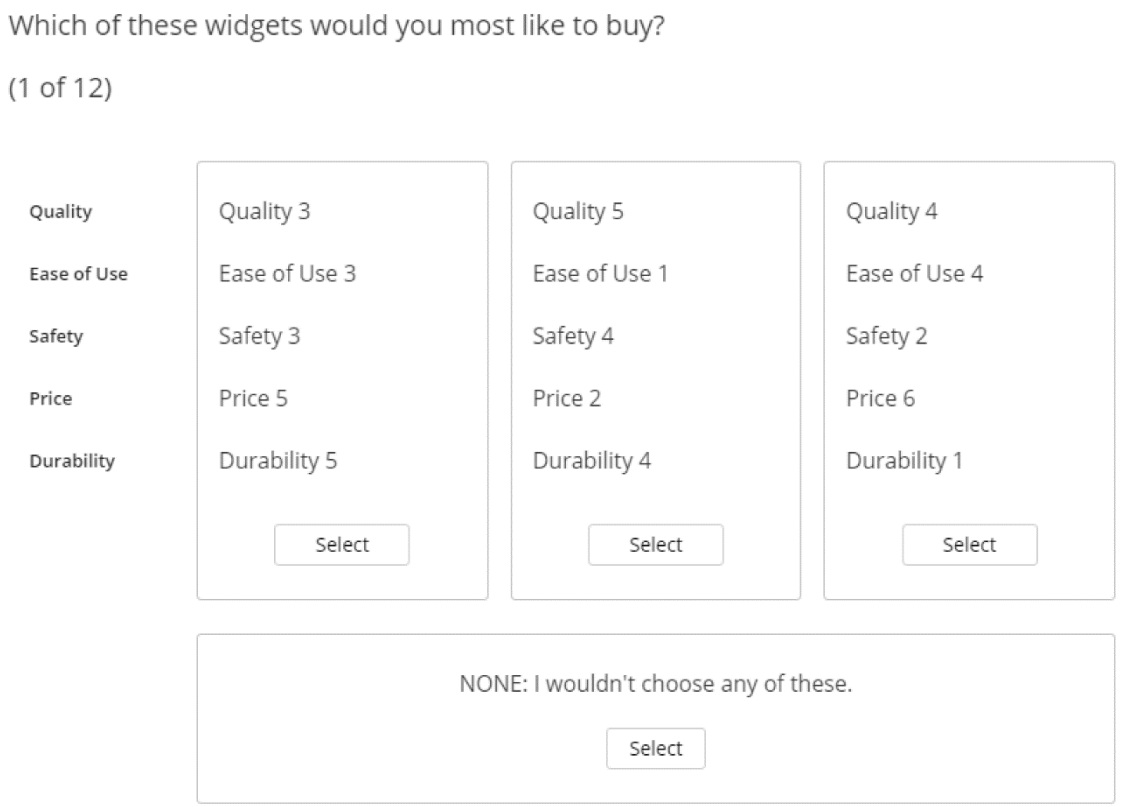
Best-Worst Case 2 Scaling as Conjoint

- With the additional question, we have a choice-based conjoint with a single product profile (in the MaxDiff question) and a “none” alternative
- You could analyze these follow-up questions to build a CBC model
- You can combine this CBC response with the MaxDiff responses and run a combined model
- In either case, you can measure interactions (e.g. if Product Quality 1 and Safety 3 work especially well together, and if together their utility exceeds the sum of their individual utilities, you can detect this, where you cannot with MaxDiff)

CONJOINT EXPERIMENT – ATTRIBUTES ARE MESSAGE ELEMENTS

Choice-Based Conjoint (CBC) Experiment

- If our message elements fall into categories, we could treat them as attributes and the individual elements as levels:



Conjoint Experiment Details

- Each respondent answers 10-15 such questions
- An experimental design controls each element, i.e. which quality, price, safety, etc. message elements appear
  - So we have utilities for each message element
  - The utilities are additive
  - If we identified significant interactions, they are also additive with the utilities
- We can build the resulting model into a simulator

Advantages of CBC

- You can prioritize elements within a category (but not across categories – for that you need Best Worst Case 2)
- You have a greater ability (more power) for testing for interactions
- Utilities are generated in the context of product preference
- We can get by with a smaller sample size (typically a minimum of 200 per separately reportable segment of customers)

Conditional Display for Natural-Looking Messages

- We may want to show combinations of attributes in a format that looks more like a message
- We can do this with conditional display
  - Instead of this . . .

    Floorplan Name: Burlwood
    # Stories: 2 stories
    # Bedrooms: 4
    Lot Size: ½ acre

  - Maybe this . . .

    The “Burlwood” floorplan is a classy two-story house with four bedrooms, situated on a ½ acre lot.”

CONJOINT EXPERIMENT – MESSAGE IS ONE OF THE ATTRIBUTES

Conjoint Experiment

- Now a complete message is one part of a product offering respondents can choose in a CBC

Assume your widget has broken and you need to buy a new one. Which of these widgets would you choose?

(1 of 12)

Feature	Good	Better	Best
Bells	1	3	10
Whistles	50	20	75
Advantage	Absent	Present	Absent
Message	Message 1	Message 2	Message 3
Price	$$	$$$	$
	Select	Select	Select

Conjoint Experiment Details

- Each respondent answers 10-15 such questions
- An experimental design controls which features, message, price appear together so that the statistical model can isolate and quantify the effect of each level of each attribute
  - So we have utilities for each level of each feature and of price
  - And we have utilities for each message
- We can build the resulting model into a simulator that shows share as a function of messaging, so we can see the sensitivity of share to messaging, or learn which message maximizes share

Advantages of CBC

- Evaluates messages in a competitive context
- Because each respondent sees 10 or so choice sets, we can get by with a smaller sample size
  - At least 400 respondents in total
  - Or at least 200 per separately reportable segment

SUMMARY

Summary

- There are lots of ways to structure experiments for messaging research
  - Monadic or sequential monadic experiments
  - MaxDiff
  - Best-Worst case 2 scaling (with or without an overall follow-up question)
  - Choice-based conjoint
- The method we choose will depend on
  - Whether we're studying message elements or entire messages
  - How many messages we want to study
  - Whether we want to know how to prioritize, bundle or combine discrete message elements