Discover-MaxDiff: How and Why It Differs from Lighthouse Studio's MaxDiff Software

Abstract and Overview

Sawtooth Software's web-based (SaaS--Software as a Service) questionnaire authoring tool is streamlined, attractive, and user-friendly. We're calling this web-based platform "Discover" and the MaxDiff component within it "Discover-MaxDiff." Discover-MaxDiff includes the essential aspects of Sawtooth Software's Lighthouse Studio MaxDiff software. This white paper describes the minor differences between MaxDiff within Discover and Lighthouse Studio in detail. Users will find Discover-MaxDiff easy to use, they'll be able to collaborate better in teams, and the results should be nearly indistinguishable from our MaxDiff package in Lighthouse Studio. All aspects, from questionnaire authoring, designing MaxDiff tasks, fielding the study, and analyzing the results are managed within the intuitive, browser-based interface.

Experimental Design Differences

- For Lighthouse Studio, the user generates and uploads a database of questionnaire versions (blocks) to the web server *prior* to inviting respondents to take the survey (default=300 versions). Entering respondents are assigned successive versions (respondent #1 receives version #1, respondent #2 receives version #2, etc.). After the 300th version has been given to a respondent, versions start again with #1.
- Discover generates questionnaire versions (blocks) on-the-fly (in the moment the respondent receives MaxDiff question #1), paying attention to achieving a high quality design for the current respondent. Different respondents receive different combinations of items to evaluate, since the algorithm is seeded differently for each respondent. A timer in the design algorithm ensures that it takes no more than about a second to generate the design, so respondents are not left waiting.

What are the ramifications of these design differences?

Both design algorithms seek to optimize 1-way, 2-way, and positional balance. Ideally, each item is shown exactly an equal number of items, each pair of items is shown in sets an equal number of times, and each item is shown in each position in the questionnaire (first, second, last, etc.) an equal number of times. But given the number of items and the number of questions (sets) per respondent, perfection on all three goals within each questionnaire version and across respondents is rarely achievable.

Lighthouse Studio generates MaxDiff designs prior to fielding the study and it achieve overall better level balance and slightly better pooled design efficiency than Discover, because Discover's design generation for the current respondent does not pay attention to previous designs given to earlier respondents. If each respondent entering a Lighthouse Studio survey ends up completing the survey, then the 1-way, 2-way, and positional balance across respondents will be better than Discover's on-the-fly approach. However, in most projects, the completion rate is usually less than 100%, so the final Lighthouse Studio designs are not quite as efficient and balanced as a preliminary analysis before collecting data would suggest.

Using robotic respondents, we have tested different MaxDiff designs[1] as typically seen in practice to compare the statistical efficiency in terms of the standard errors from aggregate logit and find that in the worst case, the loss in precision is less than 1% for Discover compared to Lighthouse Studio. ***Across our tests, standard Discover-MaxDiff designs are from 99% to 100.5% as efficient as Lighthouse Studio designs in terms of sample-level (aggregate) efficiency.***

Although it is common to measure aggregate design efficiency via aggregate logit, researchers often are more concerned with individual-level score estimation where achieving excellent 1-way level balance is very important. For challenging designs involving prohibitions, we find that Discover's on-the-fly MaxDiff designer can achieve better 1-way level balance *at the individual level* than Lighthouse Studio's MaxDiff designer. Again, Discover's designer pays attention only to the current respondent, striving for 1-way, 2-way, and positional balance within the individual. In contrast, Lighthouse Studio's MaxDiff designer jointly considers individual-level and across-respondent item balance, sometimes sacrificing individual-level balance to correct for imbalances in the aggregate 1-way and 2-way item frequency tables. For a symmetric design involving six prohibitions (where each item could be shown exactly four times to each respondent), Discover's MaxDiff designs led to perfect 1-way item balance for 300 out of 300 versions, whereas Lighthouse Studio's designer could do so for only 53 out of 300 versions. The two-way balance was better across Lighthouse Studio's designs, but interestingly enough, the aggregate design efficiency was slightly better for Discover. For a very difficult asymmetric design where it was impossible to display each level an equal times for each respondent and where challenging prohibitions were involved, we found that Discover's MaxDiff designs led to substandard 1-way level balance (compared to the best that it could have done) in 7 out of the 300 versions compared to 28 out of 300 versions for Lighthouse Studio's MaxDiff designer.

[1] Typical MaxDiff designs are those where the researcher seeks to estimate independent scores for an array of items. Standard designs may include prohibitions. They may be symmetric (it is possible for each item to appear an equal number of times for each respondent) or asymmetric (it is impossible to show each item an equal number of times for each respondent). We did encounter an exception in our testing where Discover designs were less than 99% as efficient as Lighthouse Studio designs: for a rare "best-worst case 2," (also known as best-worst conjoint, where the items and prohibitions are set up to mimic a conjoint-style design (where there are multiple attributes each having 2 or more mutually-exclusive levels). For that test, the on-the-fly Discover-Maxdiff design was 92% as efficient as the Lighthouse Studio design.

Why did Sawtooth Software decide to make Discover-MaxDiff designs on-the-fly?

- We wanted to make it extremely easy to use Discover-MaxDiff. To that aim, we decided there wasn't a compelling reason for the typical user to go through the steps of generating and examining the experimental design ahead of time. MaxDiff designs are especially robust in the face of prohibitions. As long as you ask enough questions to cover each item multiple times per respondent (and the software warns you if you don't), it's extremely hard to make a mistake.
- Not all respondents that start MaxDiff questionnaires end up finishing. Generating a perfectly balanced design ahead of time does not guarantee the final data will be perfectly balanced. On-the-fly design generation typically achieves 99% or better efficiency compared to ideal designs.
- We are targeting future software development in both Lighthouse Studio and Discover to support "constructed lists" of items for each respondent, where the items (and number of items) coming into the MaxDiff survey can be customized across respondents based on earlier answers to questions. To do this, an on-the-fly approach is required.

If the researcher wants to evaluate the efficiency and characteristics of the design before fielding the Discover-Maxdiff survey, we recommend manually creating test records, downloading the data, and examining the test experimental design. While it won't exactly be the design that results when real respondents take the survey (for example, not all respondents who start the survey will complete it and the design seed is based on the respondent number), this testing approach will let you examine the general quality and characteristics of the design across multiple respondents prior to fielding the study.

As both the Discover and Lighthouse Studio platforms mature, we will likely offer the researcher both options: either generate the design and examine its quality prior to data collection, or generate designs on-the-fly during data collection.

Utility Estimation Differences

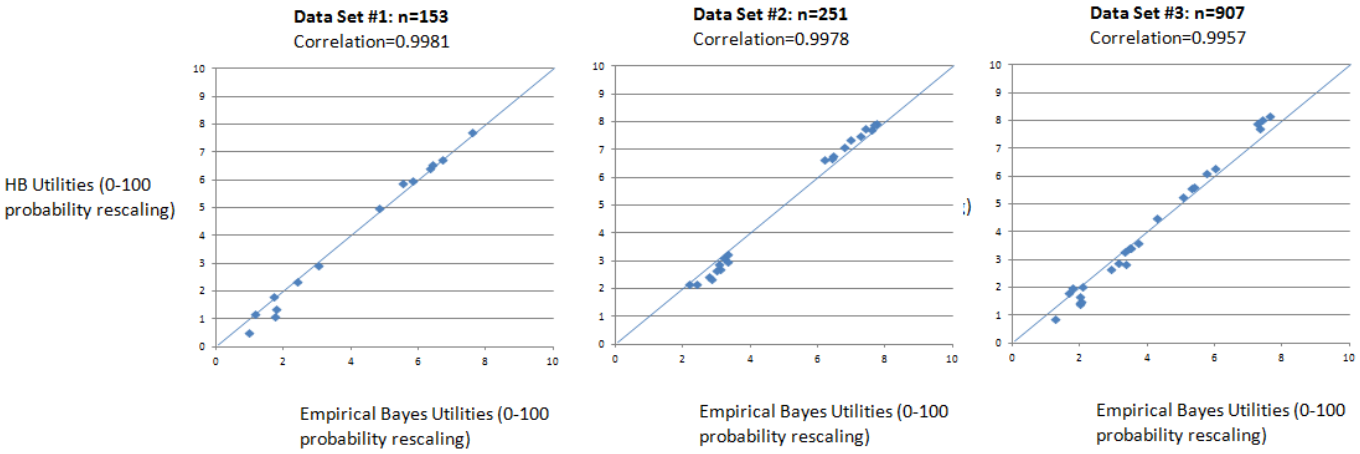
Lighthouse Studio MaxDiff uses HB (hierarchical Bayes) for individual-level score estimation. HB estimation for MaxDiff typically takes 5 minutes or longer. For Discover-Maxdiff we decided to employ a type of empirical Bayes that is *much* quicker and achieves nearly as good of results as HB (in terms of predictive validity) for studies when each item appears 2x or more per respondent (a recommended multiple). Typical datasets will take 5 to 10 seconds with empirical Bayes instead of 5 minutes or more for HB.

We examined three real MaxDiff datasets where each item was shown about 2 times per respondent. To assess the quality of the utility estimation, we randomly selected some of the MaxDiff tasks to serve as holdout observations. The individual-level hit rates for HB and empirical Bayes for these three data sets were as follows:

	HB Hit Rate	Empirical Bayes Hit Rate
Data Set #1, n=153, 18 items	72.2%	69.2%
Data Set #2, n=251, 20 items	67.5%	64.1%
Data Set #3, n=907, 24 items	60.2%	58.2%

For individual-level analysis, there are differences that favor HB estimation (by an average relative margin of 4% for the three data sets above). Our experience so far is that as the number of times each item is shown to each respondent increases, the performance gap between empirical Bayes and HB narrows.

Often researchers are more interested in the pooled (average) results across the sample, where the correlation between HB and empirical Bayes results is typically 0.99 or better. Below are the scatterplots for these same MaxDiff datasets, with empirical Bayes results on the X axis and HB on the Y-axis (the raw parameters have been exponentiated at the individual level and aggregated, so that they take on 0-100 probability scaling). The 45-degree line is shown for reference.



If the Discover-Maxdiff user is interested in using HB estimation, the data may be exported to a .CSV file that is compatible with Sawtooth Software's CBC/HB software for utility estimation. As we continue to enhance Discover-Maxdiff, we may offer HB estimation as a built-in feature.

Summary

There are some differences in MaxDiff design generation and utility estimation between Discover and Lighthouse Studio platforms, but the results tend to be quite similar. Either approach will work well in practice.

- Discover's MaxDiff is an easier and more streamlined system to use, especially for those who need a cloud-based (SaaS) platform rather than Windows.
- Discover's on-the-fly designs are typically 99% as efficient as Lighthouse Studio designs.
- Lighthouse Studio's HB estimation is about 4% better in terms of individual-level predictive hit rates than Discover's empirical Bayes approach, when each item is shown about twice per respondent. As the number of times each item is shown increases, the gap in predictive validity between HB and empirical Bayes narrows.
- The pooled (sample average) utility scores are correlated 0.99 or better when comparing Discover's empirical Bayes and Lighthouse Studio's HB estimation.