



## Funnel testing in webpage optimisation: representation, design and analysis

Heng Su & C. F. Jeff Wu

To cite this article: Heng Su & C. F. Jeff Wu (2017) Funnel testing in webpage optimisation: representation, design and analysis, *Statistical Theory and Related Fields*, 1:1, 3-14, DOI: [10.1080/24754269.2017.1332964](https://doi.org/10.1080/24754269.2017.1332964)

To link to this article: <https://doi.org/10.1080/24754269.2017.1332964>



Published online: 12 Jul 2017.



Submit your article to this journal [↗](#)



Article views: 332



View related articles [↗](#)



View Crossmark data [↗](#)



# Funnel testing in webpage optimisation: representation, design and analysis

Heng Su and C. F. Jeff Wu

School of Industrial and Systems Engineering, Georgia Institute of Technology, Wells Fargo Bank, Charlotte, NC, USA

## ABSTRACT

When optimizing webpages in order to achieve the best conversion rate, the traditional approach is to isolate and analyze them separately through a sequence of experiments. In this paper, we propose a new framework to study a system of webpages simultaneously through the use of directed graph, fractional factorial design and simple optimization algorithm. The illustrative example shows that such complicated web systems can be easily studied and optimized by using the proposed methodology.

## ARTICLE HISTORY

Received 1 March 2017  
Accepted 17 May 2017

## KEYWORDS

Conversion rate optimization;  
webgraph; factorial design;  
design of experiment

## 1. Introduction

In the information technology age, more business transactions are conducted on the Internet. Webpage has become an important source of revenue for many companies such as Amazon, Facebook, Walmart and eBay. How to design the webpages to best serve the interest of the business owners has now become a hot research topic in e-commerce (Ash, 2009). The jobs of interest completed by the visitors to the webpages are called *conversions*. Typical examples of conversions are purchases, newsletter or membership subscriptions, viewing of a page, etc. The percent of visitors completed the jobs of interest is called the *conversion rate*. A main goal of studying the webpages design is to maximise the conversion rate. This is called *conversion rate optimisation* (abbreviated as *CRO*). *CRO* has been extremely important in large IT companies in the last decade. Through this practice, companies have seen a huge increase in their profits. Moreover, companies doing *CRO* for others, such as Webtrends and SiteTuners, have grown rapidly in the last ten years and are now very popular in the IT industry. Through *CRO*, they helped the clients achieve greater business success.

Two methods are commonly used in *CRO* (Ash, 2009). The first one is called A/B test. As the name indicates, this method compares two versions of a webpage: the original version and the proposed new version. A variation of this method is called A/Bn test, where multiple proposed versions of a webpage are compared with the original design in one experiment. Hypothesis testing is used to assess the difference and the best version is chosen as the design of the webpage in the future. The second method is called multivariate test (abbreviated as *MVT*), where multiple factors, each with two or more levels, are studied in one experiment. *MVT* is usually implemented with fractional factorial designs, and

models are fitted for the conversion rates with respect to the factors. The optimisation is done by the standard method of choosing optimal level settings in design of experiment and the best combination of factor levels is used for future webpage design (Montgomery, 2012; Wu & Hamada, 2009).

A/Bn test is more commonly used in *CRO* because it is easy to understand, implement and analyse. When companies start a web campaign, they usually have different designs for their campaign page. In order to maximise the revenues from the campaign, they use the beginning part of the campaign to do an A/Bn test and use the best version as the campaign page for the rest of the campaign. On the other hand, when it comes to optimising a product page, which consists of multiple sections such as header, banner, text and pictures, *MVT* can be more efficient.

The most commonly studied page with *CRO* is the landing page, which is the first page visitors see when directed from other sources such as search engines or directly entered web addresses (Ash, 2009). Most landing pages only have the general information of the company's products and services, and conversions usually do not take place here. For example, suppose the visitors want to buy products from the website. Before they make payments, they usually have to go through the product description, view the product pictures, check the product reviews, and enter the payment information. With so many other pages involved, studying just the landing page in order to maximise the conversion rate may be an oversimplification. The series of pages the visitors went through until a possible conversion is called the *conversion funnel* (Ozolins, 2012), or abbreviated as the *funnel* if there is no ambiguity in the context. In the last example, the conversion funnel consists of the landing page, the product description page, the product picture page, the product review page and the

payment information page. Moreover, conversions may take place on different pages. For example, the visitors can subscribe the newsletters on the landing page, the product page and even on the payment confirmation page. The pages where conversions can possibly happen are called *conversion points*. If the landing page is studied with respect to only one conversion point, the results can hardly be conclusive. The set of pages associated with the conversion of interest is defined as the *conversion system*, or abbreviated as the *system* if there is no ambiguity in the context. A conversion system consists of the landing page, all the conversion points, and all other pages that link between them. For example, suppose we have three pages: the landing page, page 1 and page 2, and both page 1 and page 2 are conversion points. Suppose visitors can go from the landing page to either page 1 or page 2, and also from page 1 to page 2. Then, the conversion system consists of these three pages, and we referred this example as the *toy example*. This example will be used as the primary illustrative tool of the framework proposed in this paper. In this example, the landing page and page 1 make a conversion funnel.

In the next section, we use directed graph to represent a conversion system and use this graph representation to identify all the conversion funnels in a system. A fractional factorial design on all the pages in a system is used to conduct the experiment. In section 3, we propose an analysis strategy to optimize the conversions in a system and illustrate its application in a toy example in section 4. A simulated example is given in section 5. Concluding remarks and future research are given in section 6.

## 2. Representation and design

The idea of using directed graph to study the Internet originates from a concept called webgraph (Donato, Laura, Leonardi, & Millozzi, 2004), where webpages are viewed as vertices and the linkage relationships between pages are expressed with directed edges in the graph. In computer science, the size of the webgraph being studied is usually very large (over millions), and researchers are interested in the large-scale properties such as in/out distributions, connectivity and cyclic patterns. The results are used to identify communities and hubs, filter spams, rank pages and predict the growth of the Internet (Donato et al., 2004). In this work, we use this idea, and represent a conversion system with directed graph. Here we are most interested in how to identify all the conversion funnels from the graphical structure.

The representation is straightforward. All the webpages in a conversion system are viewed as vertices of the graph. If there is a hyperlink on page X referring to page Y, draw a directed edge from X to Y. For the toy example, the conversion system consists of three pages:

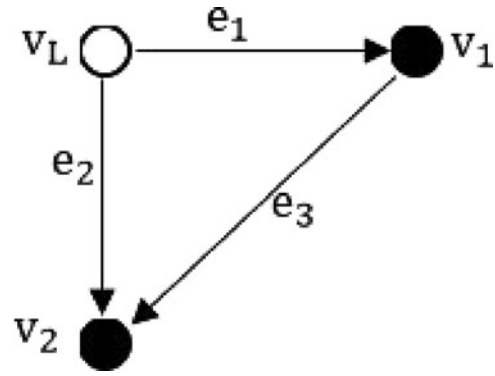


Figure 1. Directed graph representation of the toy example.

the landing page, page 1 and page 2, and we denote these three vertices as  $v_L$ ,  $v_1$  and  $v_2$ , respectively. On the landing page, there are links referring to both page 1 and page 2. Therefore, there are two edges starting at  $v_L$  and directing to  $v_1$  and  $v_2$ , respectively. We denote them as  $e_1$  and  $e_2$  correspondingly. Furthermore, there is a link on page 1 referring to page 2. Draw another edge from  $v_1$  to  $v_2$  denoted by  $e_3$ . There are no more links in this conversion system, so the toy example is represented by the following graph in Figure 1. The conversion points are marked in solid dots in contrast to others.

After representing the conversion system as a directed graph, the next step is to identify all the conversion funnels. By definition, a conversion funnel is a series of pages that a visitor has gone through before making a possible conversion. Since visitors always start with the landing page, and make conversions on the conversion points, in the graph, a conversion funnel is a path from the landing page to a conversion point. To identify all the conversion funnels in the system amounts to finding all the paths connecting the landing page and all the conversion points. In the toy example, page 1 and page 2 are both conversion points. For page 2, there are two paths connecting it to the landing page:  $v_L$  via  $e_1$  to  $v_1$  then via  $e_3$  to  $v_2$ , and  $v_L$  via  $e_2$  to  $v_2$ . These two conversion funnels are referred as  $CF_1$  and  $CF_2$ . For page 1, there is only *one* path connecting it to the landing page:  $v_L$  via  $e_1$  to  $v_1$ . This conversion funnel is referred to as  $CF_3$ . Since there are no more conversion points in this system, the toy example has three different conversion funnels.

Before this work, some researchers have realised the concept of conversion funnels and have done experiments using this concept (Qualaroo, 2014). However, they used one-page-at-a-time method, and studied the pages in a funnel sequentially. Note that, even for a small conversion system like the toy example, there are already three different conversion funnels. The old method is extremely time-consuming and ignores any interactions between different pages. In this work, we study the conversion system as a whole and design one experiment for *all* the pages involved.

Since there are multiple pages involved in the experiment, A/Bn test will be inefficient. How about the MVT? In MVT, the first step is to identify the factors being studied. Since this experiment considers all the pages in the conversion system at one time, the set of factors consists of the factors from all the pages. For example, in the toy example, suppose each page has two factors to be studied: A and B from the landing page, C and D from page 1 and E and F from page 2. The factors being studied in this experiment are A, B, C, D, E and F. After identifying the correct set of factors, the next step is to construct a fractional factorial design for them. For details on the choice of designs, the reader may refer the book by Wu and Hamada (2009).

For ease of implementation, we assume each factor has two levels. In this case, a  $2^{k-p}$  design is used, where  $k$  factors are being studied, each at two levels denoted by + and -. It is the  $p^{-1}$  fraction of the  $2^k$  full factorial design. In some other cases, mixed-level designs are also used.

The conversion data are collected as binary responses: 1 means conversion made and 0 means no conversion. The conversions for each funnel are recorded separately for ease of analysis. Take the toy example, if a conversion is made page 2, and the visitor comes directly from the landing page, this conversion is recorded as data with  $CF_2$ .

### 3. Analysis

Define the total conversion rate as the weighted sum of the conversion rates of all the conversion funnels in the system, where the weights reflect the importance of the funnels to the business owners. For example, in studying the sales of shoes, the weights can be the price of shoes in different funnels. When considering the subscription rate of the newsletters, the weights can be set to be equal. The goal is then to maximize the total conversion rate.

Take the toy example for illustration. We have 8 factors to be analyzed: A through F plus  $E'$  and  $F'$  for the additional variant of page 2 in the conversion funnel 2. The reason to add  $E'$  and  $F'$  is that the optimal settings for page 2 under conversion funnel 1 may not be optimal for conversion funnel 2. Suppose you want to design a three-page conversion system to maximize the membership subscriptions. For simplicity, assume this system is the same as in the toy example. Then, let us check which page(s) can have multiple versions and why. The landing page appears in all three conversion funnels, but the visitors have to start with the landing page, so all three funnels share the same version of the landing page. Page 1 appears in  $CF_1$  and  $CF_3$ , but visitors are always directed from the landing page to page 1. Therefore, page 1's in the two funnels are the same. Page 2 appears in  $CF_1$  and  $CF_2$ , but before page 2, visitors have viewed different pages for these two funnels:

for  $CF_1$ , visitors have viewed the landing page as well as page 1 before page 2, whereas for  $CF_2$ , visitors have only viewed the landing page before page 2. Since page 1 is not in  $CF_2$ , visitors in the two funnels have been exposed to different information before page 2. Therefore, page 2's in  $CF_1$  and  $CF_2$  can have different versions. This can be explained with the following example. Suppose page 1 shows the benefits of the membership and page 2 shows the price in the original design. Normal visitors check the benefits and compare them with the price on page 2 and determine whether to subscribe the membership. If the visitor does not care about price, she/he may subscribe after viewing the benefits, and if the visitor is eager to know the price, she/he may jump to page 2 directly from the landing page. For the last visitor, she/he does not have the information about the benefits; the decisions made might be biased. Suppose the membership price is high, but also offers great benefits whose monetary value may exceed the cost. For normal visitors, after comparing the cost with benefits, they are likely to pay for the membership price. But for visitors that have skipped the benefits page, just by looking at the price, they might think it is too high and decide not to go for it. These two decisions are made with different amounts of information, and the second one is biased. With the concept of the conversion funnel, it is immediately noticed that these two decisions are made on two different funnels. In order to correct the second situation, we simply show the visitors another version of page 2, which has the benefits as well as the price of the membership, if they come directly from the landing page.

The idea of the analysis is to study the conversion rate of a specific conversion funnel instead of specific pages, since a page can be related to multiple conversion funnels. For example, in the toy example, page 2 is the conversion point for  $CF_1$ , but it is also the conversion point for  $CF_2$ . For each conversion funnel, we identify its related factors from its constitution of web pages. As in the toy example,  $CF_1$  is related to page 1 and 2, so the model for  $CF_1$  would have A, B, C and D as the candidate factors. Similarly, we can identify the candidate factors for  $CF_2$  and  $CF_3$ . After all the candidate factors are identified for all conversion funnels, we can then build linear models individually for each of them. For information on the analysis of fractional factorial design, the readers can review the book by Wu and Hamada (2009).

The total conversion rate is obtained through a linear combination of the conversion rates for individual conversion funnels. The optimization is then done with respect to the 8 factors including  $E'$  and  $F'$  as discussed in the previous paragraphs.

### 4. Illustration of analysis with the toy example

In this section, we use the toy example and simulate a set of data to illustrate the analysis strategy. Recall that the

**Table 1.** Design matrix, toy example.

A	B	C	D	E	F
1	1	1	1	1	1
1	1	1	-1	1	-1
1	1	-1	1	-1	-1
1	1	-1	-1	-1	1
1	-1	1	1	-1	-1
1	-1	1	-1	-1	1
1	-1	-1	1	1	1
1	-1	-1	-1	1	-1
-1	1	1	1	-1	1
-1	1	1	-1	-1	-1
-1	1	-1	1	1	-1
-1	1	-1	-1	1	1
-1	-1	1	1	1	-1
-1	-1	1	-1	1	1
-1	-1	-1	1	-1	1
-1	-1	-1	-1	-1	-1

conversion system in the toy example consists of three pages: the landing page, page 1 and page 2, and there are three conversion funnels in the system:  $CF_1$ ,  $CF_2$  and  $CF_3$ . Note that we suppose each page has two factors to be studied, so a  $2^{6-2}$  design is used for the simulation, where each row of the design matrix represents a version of the conversion system. The design matrix is given in Table 1, whose defining relations are  $I = ABCE = BCDF = ADEF$ .

For each simulation, we first choose a version of the conversion system from the 16 candidates in the design table with equal probability, and then simulate the visitors' behaviour in the chosen system. The visitor always starts with the landing page, and she/he then have three choices: go to page 1, go to page 2 or leave the system. The first step is to simulate the decision on the landing page. Suppose the visitor goes to page 1 or page 2 with probabilities  $t_1$  or  $t_2$ , respectively, where  $t_1$  and  $t_2$  are functions of the factors related to the landing page, i.e., A and B. All the functions of decision probabilities used in this simulation are listed in Table 2. The decisions are made sequentially. First, we determine if the visitor goes to page 1. If she/he does not go to page 1, then we check if she/he goes to page 2. If the choice of the second one is still negative, she/he leaves the system. So to be specific,  $t_2$  is the probability of the visitor going to page 2 given that she/he does not go to page 1. If the visitor chooses to go to page 1, then on page 1, there are still three choices: make a conversion, go to page 2 or leave. We then simulate his/her decision on page 1. Suppose she/he makes a conversion with probability  $c_1$ , If she/he does not make a conversion, then goes to page

**Table 2.** Functions for decision probabilities.

$t_1 = 0.25 - 0.1A$
$t_2 = 0.42 + 0.21B$
$t_3 = 0.1 + 0.08A - 0.3D + 0.1AD$
$c_1 = 0.38 - 0.05B + 0.12C + 0.08BD$
$c_2 = 0.07 + 0.2E + 0.07AE$
$c_3 = 0.15 + 0.06D - 0.1F + 0.02AD + 0.01CF$

2 with probability  $t_3$ . If the decisions are both negative, she/he leave the system. Note that, although the choices are made on page 1, the two probabilities  $c_1$  and  $t_3$  are functions of factors related to both the landing page and page 1, i.e., A, B, C and D, because it is believed that the information on the landing page will affect the visitors' behaviour thereafter. Finally, for visitors who land on page 2, there are two types: they come directly from the landing page, or have visited page 1. This represents the two conversion funnels  $CF_2$  and  $CF_1$ , respectively. Since page 2 has no link referring to others, the choices are whether to make conversions. The decisions are made as follows. If the visitor comes via  $CF_2$ , she/he converts with probability  $c_2$ ; otherwise, convert with probability  $c_3$ . Note that  $CF_2$  consists of two pages. Therefore,  $c_2$  is a function of the factors related to those two pages, i.e., A, B, E and F. Similarly,  $c_3$  is a function of factors A, B, C, D, E and F. Each simulation terminates when the visitor either makes a conversion or leaves the system. We ran 10,000 simulations, and recorded the conversions of each funnel separately. The conversion rates are given in Table 3.

For simplicity, assume all the conversion funnels have equal weights in the toy example. Denote the total conversion rate as  $CR_T$ , and the conversion rates for  $CF_1$ ,  $CF_2$  and  $CF_3$  as  $CR_1$ ,  $CR_2$  and  $CR_3$  respectively. The objective function for optimisation can be written as

$$CR_T = CR_1 + CR_2 + CR_3.$$

Then we build models for  $CR_1$ ,  $CR_2$  and  $CR_3$  separately. Recall that the first step in modelling the conversion rate of a funnel is to identify the related factors. For  $CR_1$ , since  $CF_1$  consists of three pages, the landing page, page 1 and page 2, the factors considered in this model are A, B, C, D, E, and F. Similarly, for  $CR_2$  and  $CR_3$ , the factors considered are A, B, E and F and A, B, C and D, respectively.

The model building is then straightforward. For  $CR_1$ , it is a function of all the six factors. The corresponding design matrix and responses considered in this model

**Table 3.** Conversion rates for different funnels, toy example.

$CF_1$	$CF_2$	$CF_3$
0.00000000	0.16349206	0.08730159
0.00960000	0.16960000	0.05440000
0.00000000	0.00000000	0.05546218
0.00000000	0.00000000	0.02073365
0.00000000	0.00000000	0.06833333
0.00000000	0.00000000	0.10194175
0.00000000	0.06718750	0.03593750
0.01093750	0.06406250	0.07812500
0.00000000	0.00000000	0.14803150
0.02044025	0.00000000	0.12421384
0.00000000	0.07833333	0.09166667
0.00000000	0.08258258	0.05255255
0.00000000	0.02508361	0.16722408
0.00000000	0.03475513	0.21169036
0.00000000	0.00000000	0.07954545
0.01716069	0.00000000	0.14820593



**Table 4.** Design matrix and response data, conversion rates for  $CF_1$ .

A	B	C	D	E	F	$CR_1$
1	1	1	1	1	1	0.00000000
1	1	1	-1	1	-1	0.00960000
1	1	-1	1	-1	-1	0.00000000
1	1	-1	-1	-1	1	0.00000000
1	-1	1	1	-1	-1	0.00000000
1	-1	1	-1	-1	1	0.00000000
1	-1	-1	1	1	1	0.00000000
1	-1	-1	-1	1	-1	0.01093750
-1	1	1	1	-1	1	0.00000000
-1	1	1	-1	-1	-1	0.02044025
-1	1	-1	1	1	-1	0.00000000
-1	1	-1	-1	1	1	0.00000000
-1	-1	1	1	1	-1	0.00000000
-1	-1	1	-1	1	1	0.00000000
-1	-1	-1	1	-1	1	0.00000000
-1	-1	-1	-1	-1	-1	0.01716069

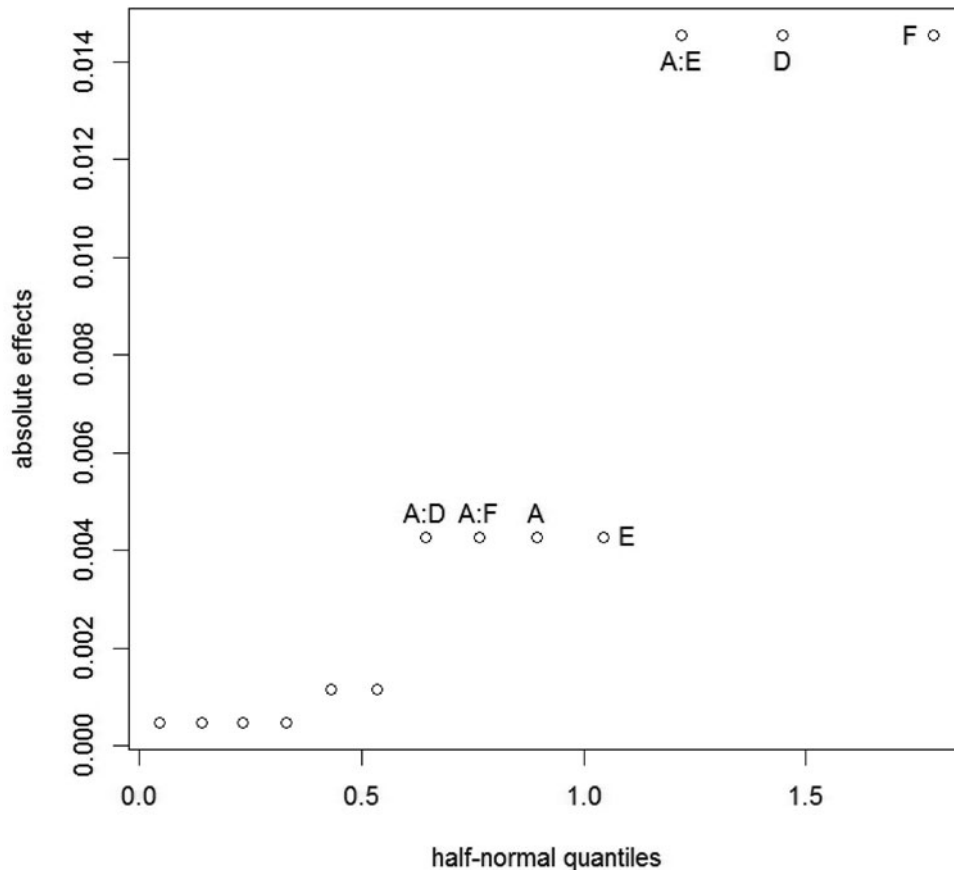
are shown in Table 4. The first step of modelling is to make a half-normal plot to identify significant effects. In Figure 2, it is clearly seen that main effects F and D and two-factor interaction (abbreviated as 2fi thereafter) AE are the most significant. Because these three terms have the same magnitude, we denote them as group 1. They are followed by main effects E and A and 2fi's AF and D, whose magnitudes are the same, and denoted as group 2. The other effects are not significant. Therefore, the model for  $CR_1$  has seven terms. The  $R^2$

value for this model is 99.12% and the  $p$  values for the significant effects in group 1 and group 2 are  $1.91e-7$ , and 0.00133, respectively. The explicit expression of the model is

$$\begin{aligned}
 CR_1 = & 0.00336337 - 0.00336337F - 0.00336337D \\
 & - 0.0010665E - 0.0010665A + 0.00336337AE \\
 & + 0.0010665AF + 0.0010665AD. \tag{1}
 \end{aligned}$$

For  $CR_2$ , the corresponding funnel consists of two pages: the landing page and page 2. Therefore, only factors A, B, E and F are considered. The corresponding design matrix and responses used in this model are shown in Table 5. We start the analysis by drawing a half-normal plot to identify significant effects. In Figure 3, it is clearly seen that the main effect E is the most significant. It is followed by B and BE, and then A and AE. The other effects are not significant. Therefore, the model for  $CR_2$  has five terms. The  $R^2$  value for this model is 97.28% and the  $p$  values for the significant effects are  $4.11e-8$ ,  $6.71e-5$ ,  $6.71e-5$ , 0.0377% and 0.0377%, respectively. The explicit model is written below:

$$\begin{aligned}
 CR_2 = & 0.042819 + 0.042819E + 0.018932B \\
 & + 0.015224A + 0.018932BE + 0.015224AE. \tag{2}
 \end{aligned}$$



**Figure 2.** Half-normal plot, conversion rate for  $CF_1$ .

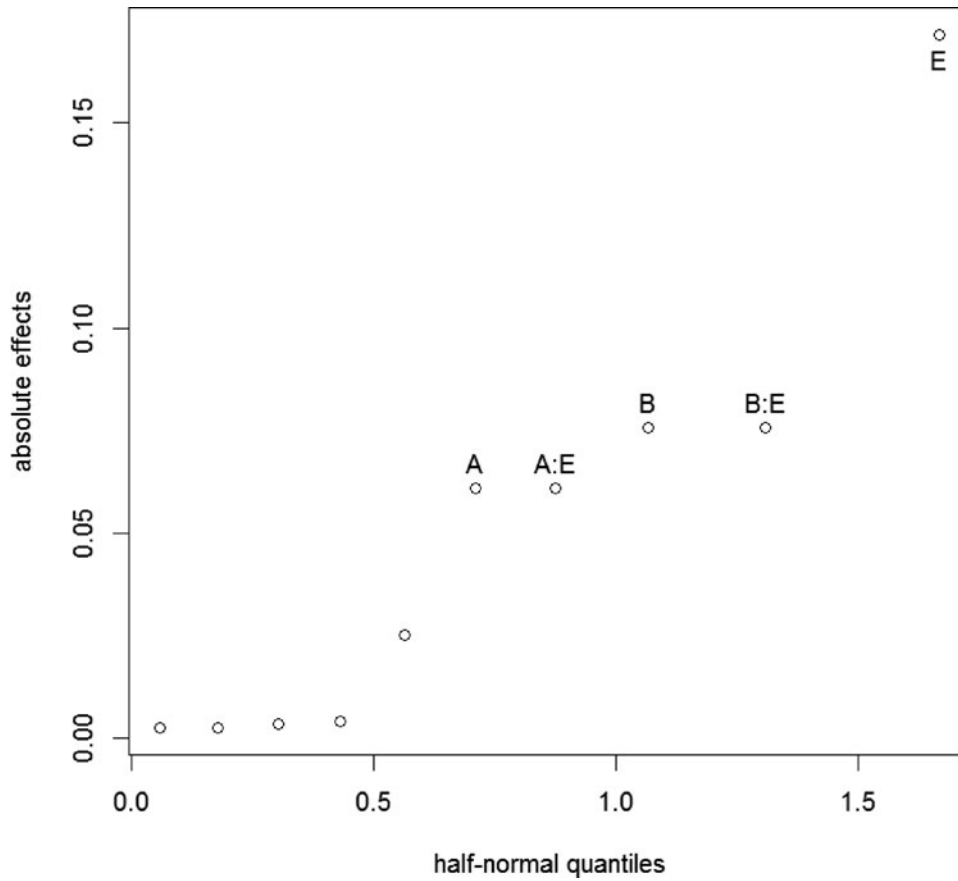


Figure 3. Half-normal plot, conversion rate for  $CF_2$ .

Table 5. Design matrix and response data, conversion rates for  $CF_2$ .

A	B	E	F	$CR_2$
1	1	1	1	0.16349206
1	1	1	-1	0.16960000
1	1	-1	-1	0.00000000
1	1	-1	1	0.00000000
1	-1	-1	-1	0.00000000
1	-1	-1	1	0.00000000
1	-1	1	1	0.06718750
1	-1	1	-1	0.06406250
-1	1	-1	1	0.00000000
-1	1	-1	-1	0.00000000
-1	1	1	-1	0.07833333
-1	1	1	1	0.08258258
-1	-1	1	-1	0.02508361
-1	-1	1	1	0.03475513
-1	-1	-1	1	0.00000000
-1	-1	-1	-1	0.00000000

Table 6. Design matrix and response data, conversion rates for  $CF_3$ .

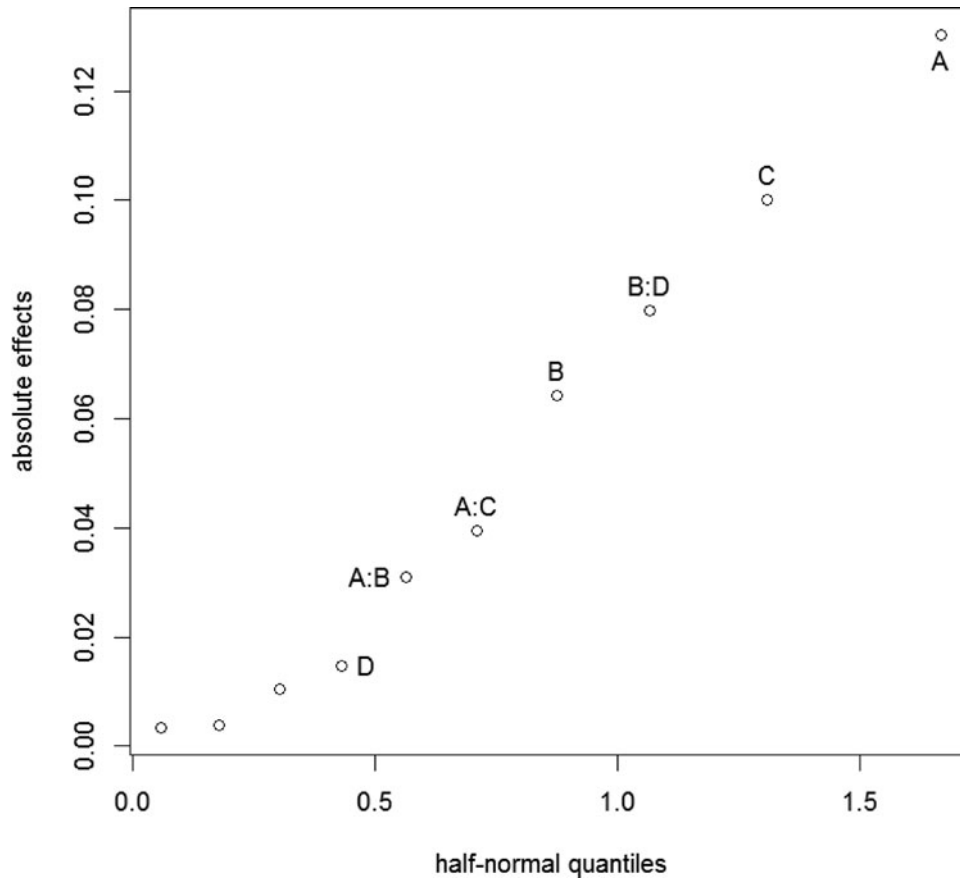
A	B	C	D	$CR_2$
1	1	1	1	0.08730159
1	1	1	-1	0.05440000
1	1	-1	1	0.05546218
1	1	-1	-1	0.02073365
1	-1	1	1	0.06833333
1	-1	1	-1	0.10194175
1	-1	-1	1	0.03593750
1	-1	-1	-1	0.07812500
-1	1	1	1	0.14803150
-1	1	1	-1	0.12421384
-1	1	-1	1	0.09166667
-1	1	-1	-1	0.05255255
-1	-1	1	1	0.16722408
-1	-1	1	-1	0.21169036
-1	-1	-1	1	0.07954545
-1	-1	-1	-1	0.14820593

Finally, for  $CR_3$ , the corresponding conversion funnel  $CF_3$  consists of two pages: the landing page and page 1. Therefore, factors A, B, C and D should be considered. The corresponding design matrix and responses used in this model are shown in Table 6. We again start the analysis by drawing a half-normal plot to identify significant effects. In Figure 4, it is seen that the main effect A is the most significant. It is followed by C, BD, B, AC and AB. The other effects are not significant. Therefore, the model for  $CR_3$  has six terms. The  $R^2$  value for this model is 98.3% and the  $p$  values for significant

effec are  $1.23e-7$ ,  $1.16e-6$ ,  $7.72e-6$ ,  $4.47e-5$ , 1.51% and 6.42%, respectively. The explicit model expression is given below:

$$\begin{aligned}
 CR_3 = & 0.095335 - 0.032556A + 0.025057C \\
 & - 0.01604B + 0.19968BD \\
 & - 0.009842AC + 0.007735AB. \quad (3)
 \end{aligned}$$

Recall that page 2 appears in both  $CF_1$  and  $CF_2$  and can have two different versions. Replace the factors E



**Figure 4.** Half-normal plot, conversion rate for  $CF_3$ .

and F in (2) with  $E'$  and  $F'$ , respectively. Now, put all three models together, we have the total conversion rate expressed as a function of all the eight factors:

$$\begin{aligned}
 CR_T = & 0.1415174 - 0.0183985A + 0.002892B \\
 & + 0.025057C - 0.0036337D - 0.0010665E \\
 & - 0.0036337F + 0.042819E' + 0.007735AB \\
 & - 0.009842AC + 0.0010665AD \\
 & + 0.0036337AE + 0.0010665AF + 0.015224AE' \\
 & + 0.019968BD + 0.018932BE'.
 \end{aligned}$$

To maximise  $CR_T$ , we find the optimal level settings of the eight factors. By checking all the possible combinations of the eight factors, it is seen that by setting A, E and F to  $-$ , and B, C, D and  $E'$  to  $+$ , we have the maximal expected conversion rate of this system, which is 26.12%. Because  $F'$  does not appear in any significant terms, we can choose either setting according to other considerations.

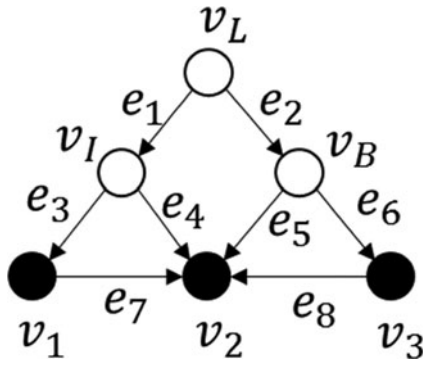
Since the objective of this methodology is to maximize the overall conversion rate, one might suggest using the overall conversion rate as the response and fitting the model with all available factors. In the beginning, this approach may seem simple and obvious with only a few webpages. However, as the size of the web system increases, the number of factors can grow exponentially. Fitting a model with so many variables may be

infeasible. Instead, if we study each conversion funnel separately, the difficulty of model fitting within the conversion funnel can be well managed. On the other hand, this bottom-up approach gives the user more flexibility: would the weight for the conversion funnel changes per user perception changes, the proposed method can easily be adapted to this situation by changing the corresponding weight in the total conversion rate formula instead of fitting a new model. In addition, the user may also gain a deeper understanding of each of the conversion funnels: the user would be able to identify factors that could affect a particular conversion funnel the most and study deeper into those factors for this conversion funnel if needed.

## 5. Simulated example

In this section, we will demonstrate the idea of funnel testing with a more complicated example. Consider a conversion system that consists of six pages. The first page is the landing page ( $v_L$ ). All customers start visiting the conversion system with this page. The second page is the individual page ( $v_1$ ). It is the page showing information for individual customers. The third page is the business page ( $v_B$ ). It is the page showing information for business customers. There are three more pages, called Product 1 page ( $v_1$ ), Product 2 page ( $v_2$ )





**Figure 5.** Directed graph representation of the simulated example.

and Product 3 page ( $v_3$ ) respectively, for customers to make conversions for three different kinds of products. The linkage relationship between pages can be described as follows. The customers always start with the landing page, where they have three choices: go to individual page, go to business page, or leave the system. If the customer goes to individual page, she/he then has three choices: go to Product 1 page, go to Product 2 page, or leave the system. Similarly, if the customer goes to business page, she/he then also has three choices: go to Product 3 page, go to Product 2 page, or leave the system. Customers can convert on any of the Product pages, or leave the system. For customers on Product 1 page or Product 3 page, they have one more choice to go to Product 2 page.

We start our analysis by representing this conversion system with the directed graph in Figure 5. The six pages are viewed as six vertices  $v_L$ ,  $v_I$ ,  $v_B$ ,  $v_1$ ,  $v_2$ , and  $v_3$ , respectively, and the linkage relationships are viewed as eight directed edges denoted by  $e_1, \dots, e_8$ .  $v_1$ ,  $v_2$  and  $v_3$  are marked in solid dots in contrast to others, meaning that they are conversion points.

The next step is to identify all the conversion funnels from the directed graph representation. By definition, a conversion funnel is a path from the landing page to a conversion point. For the conversion point  $v_1$ , only the path  $v_L \rightarrow v_I \rightarrow v_1$  connects  $v_L$  to it. We denote this conversion funnel by  $CF_1$ . Similarly, only one path  $v_L \rightarrow v_B \rightarrow v_3$  connects  $v_L$  to  $v_3$ , which is denoted by  $CF_3$ . For the last conversion point  $v_2$ , there are four paths connecting  $v_L$  to it:  $v_L \rightarrow v_I \rightarrow v_2$ ,  $v_L \rightarrow v_B \rightarrow v_2$ ,  $v_L \rightarrow v_I \rightarrow v_1 \rightarrow v_2$ , and  $v_L \rightarrow v_B \rightarrow v_3 \rightarrow v_2$ , which are represented by  $CF_{2I}$ ,  $CF_{2B}$ ,  $CF_{21}$  and  $CF_{23}$ , respectively.

Suppose each page has two factors to be investigated (A and B for the landing page, C and D for individual page, E and F for business page, G and H for Product 1 page, I and J for Product 2 page, and K and L for Product 3 page), and each factor has two levels. A  $2_{IV}^{12-6}$  design is used for the experiment. The design matrix is shown in Table 7. The defining relations of this design is  $I = ABCG = ABDH = ACDEI = ACDFJ =$

$ABEFK = BCEDFL$ . Each row of the matrix represents a version of the conversion system to be studied.

Take the landing page and its corresponding factors A and B for example. Factor A may represent the choice of the header, which has two candidates with version 1 denoted by level + and version 2 denoted by -. Similarly, B can be the choice of the main picture on the landing page, with + being version 1 and - being version 2. Other factors of other pages can be interpreted similarly. In general, each factor represents one element of its corresponding page that we want to investigate. In this example, each element has two candidate versions and we want to decide which one is better. As discussed in Section 1, elements can be headers, banners, texts, pictures, etc.

The experiment is done by simulations. For each simulation, we first choose a version of the conversion system from the 64 candidates in the design table with equal probability, and then simulate customers' behaviour in the chosen system. The customers always start with the landing page. Recall that, on the landing page, they have three choices: go to individual page, go to business page, or leave the system. We simulate their decisions in the following way: they go to individual page with probability  $t_{LI}$ , where  $t_{LI}$  represents the transition probability from the landing page to individual page; they go to business page with probability  $t_{LB}$ . If the simulated decisions of the above two statements are both negative, customers leave the system. If the simulated decisions are both positive for the two statements, the customers go to either page with equal probability. For customers on individual page, we simulate their decisions among the three choices in a similar way: they go to Product 1 page with probability  $t_{11}$ , and go to Product 2 page with probability  $t_{12}$ . Similarly, for customers on business page, we suppose they go to Product 3 page with probability  $t_{B3}$ , and go to Product 2 page with probability  $t_{B2}$ . For customers on either Product 1 page or Product 3 page, they make conversions with probability  $c_1$  or  $c_3$ , respectively. If they do not make conversions, they can further go to Product 2 page with probability  $t_{12}$  or  $t_{32}$  respectively. Otherwise, they leave the system. Finally, for customers on Product 2 page, they can either make conversions or leave the system. The probabilities for them to make conversions are  $c_{2I}$ ,  $c_{2B}$ ,  $c_{21}$ , and  $c_{23}$ , respectively, depending on the conversion funnels they are from. The simulation is terminated when the customers leave the system or a conversion is made. All the decision probabilities used in the simulations are listed in Table 8. If the calculated probabilities are less than zero, we suppose they are zero in the simulation. The choice of the decision probabilities is somewhat arbitrary but with the following rationalisation. Take  $t_{LI}$  for example, we have  $t_{LI} = 0.5 + 0.1A$ , which is a function of the main effect A of the landing page. According to the assumption, A has two levels denoted by + and -. Therefore, the

**Table 7.** Design matrix, simulated example.

Run	A	B	C	D	E	F	G	H	I	J	K	L
1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	-1	1	1	1	-1	-1	-1
3	1	1	1	1	-1	1	1	1	-1	1	-1	-1
4	1	1	1	1	-1	-1	1	1	-1	-1	1	1
5	1	1	1	-1	1	1	1	-1	-1	-1	1	-1
6	1	1	1	-1	1	-1	1	-1	-1	1	-1	1
7	1	1	1	-1	-1	1	1	-1	1	-1	-1	1
8	1	1	1	-1	-1	-1	1	-1	1	1	1	-1
9	1	1	-1	1	1	1	-1	1	-1	-1	1	-1
10	1	1	-1	1	1	-1	-1	1	-1	1	-1	1
11	1	1	-1	1	-1	1	-1	1	1	-1	-1	1
12	1	1	-1	1	-1	-1	-1	1	1	1	1	-1
13	1	1	-1	-1	1	1	-1	-1	1	1	1	1
14	1	1	-1	-1	1	-1	-1	-1	1	-1	-1	-1
15	1	1	-1	-1	-1	1	-1	-1	-1	1	-1	-1
16	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1
17	1	-1	1	1	1	1	-1	-1	1	1	-1	-1
18	1	-1	1	1	1	-1	-1	-1	1	-1	1	1
19	1	-1	1	1	-1	1	-1	-1	-1	1	1	1
20	1	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1
21	1	-1	1	-1	1	1	-1	1	-1	-1	-1	1
22	1	-1	1	-1	1	-1	-1	1	-1	1	1	-1
23	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1
24	1	-1	1	-1	-1	-1	-1	1	1	1	-1	1
25	1	-1	-1	1	1	1	1	-1	-1	-1	-1	1
26	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1
27	1	-1	-1	1	-1	1	1	-1	1	-1	1	-1
28	1	-1	-1	1	-1	-1	1	-1	1	1	-1	1
29	1	-1	-1	-1	1	1	1	1	1	1	-1	-1
30	1	-1	-1	-1	1	-1	1	1	1	-1	1	1
31	1	-1	-1	-1	-1	1	1	1	-1	1	1	1
32	1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1
33	-1	1	1	1	1	1	-1	-1	-1	-1	-1	1
34	-1	1	1	1	1	-1	-1	-1	-1	1	1	-1
35	-1	1	1	1	-1	1	-1	-1	-1	1	-1	-1
36	-1	1	1	1	-1	-1	-1	-1	1	1	-1	1
37	-1	1	1	-1	1	1	-1	1	1	1	-1	-1
38	-1	1	1	-1	1	-1	-1	1	1	-1	1	1
39	-1	1	1	-1	-1	1	-1	1	-1	1	1	1
40	-1	1	1	-1	-1	-1	-1	1	-1	-1	-1	-1
41	-1	1	-1	1	1	1	1	-1	1	1	-1	-1
42	-1	1	-1	1	1	-1	1	-1	1	-1	1	1
43	-1	1	-1	1	-1	1	1	-1	-1	-1	1	1
44	-1	1	-1	1	-1	-1	1	-1	-1	-1	-1	-1
45	-1	1	-1	-1	1	1	1	1	-1	-1	-1	1
46	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1
47	-1	1	-1	-1	-1	1	1	1	1	1	-1	1
48	-1	1	-1	-1	-1	-1	1	1	1	1	-1	1
49	-1	-1	1	1	1	1	1	1	-1	-1	1	-1
50	-1	-1	1	1	1	-1	1	1	-1	1	-1	1
51	-1	-1	1	1	-1	1	1	1	1	-1	-1	1
52	-1	-1	1	1	-1	-1	1	1	1	1	1	-1
53	-1	-1	1	-1	1	1	1	-1	1	1	1	1
54	-1	-1	1	-1	1	-1	1	-1	1	-1	-1	-1
55	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1
56	-1	-1	1	-1	-1	-1	1	-1	-1	-1	1	1
57	-1	-1	-1	1	1	1	-1	1	1	1	1	1
58	-1	-1	-1	1	1	-1	-1	1	1	-1	-1	-1
59	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	-1
60	-1	-1	-1	1	-1	-1	-1	1	-1	-1	1	1
61	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	-1
62	-1	-1	-1	-1	1	-1	-1	-1	-1	1	-1	1
63	-1	-1	-1	-1	-1	1	-1	-1	1	-1	-1	1
64	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	-1

transition probability from the landing page to individual page is  $0.6 (= 0.5 + 0.1)$  if factor A is set to level +, and  $0.4 (= 0.5 - 0.1)$  if A is set to -. Similarly, for  $t_{12}$ , we have  $t_{12} = 0.05 - 0.05DH + 0.05C - 0.05AG$ . It is a function of the main effect C and 2fi's DH and AG, which involves five factors A, C, D, G and H. Among the five factors, A is from the landing page, C and D are

from individual page, and G and H are from Product 1 page. Note that  $t_{12}$  is the transition probability from Product 1 page to Product 2 page, which is part of  $CF_{21}$ . Recall that  $CF_{21}$  is the path  $v_L \rightarrow v_1 \rightarrow v_1 \rightarrow v_2$ . Before making this decision, customers have gone through the landing page, individual page and Product 1 page along the path. Therefore, it is assumed that only factors on

**Table 8.** Functions for decision probabilities.

$$\begin{aligned}
 t_{LJ} &= 0.5 + 0.1A \\
 t_{LB} &= 0.2 - 0.1B \\
 t_{H1} &= 0.1 + 0.1A - 0.1C + 0.1BC \\
 t_{J2} &= 0.05D - 0.05AD \\
 t_{B3} &= 0.1 + 0.1E - 0.1F + 0.1C - 0.1EF \\
 t_{B2} &= 0.1A - 0.1AE \\
 t_{12} &= 0.05 - 0.05DH + 0.05C - 0.05AG \\
 t_{32} &= 0.05K - 0.05EL \\
 c_1 &= 0.1 + 0.1G - 0.1CG \\
 c_3 &= 0.1 + 0.1K - 0.1FL + 0.05BK - 0.05AF \\
 c_{2I} &= 0.1 + 0.1I - 0.1J - 0.1CJ + 0.1BI \\
 c_{2B} &= 0.1 + 0.1J - 0.1IJ + 0.05E - 0.05AJ \\
 c_{21} &= 0.05I - 0.05DJ + 0.05H + 0.05CI \\
 c_{23} &= 0.05J - 0.05E - 0.05FK + 0.1IJ
 \end{aligned}$$

these three pages can affect this decision. The 2fi DH in  $t_{12}$  can be interpreted as follows: factor D on individual page and factor H on Product 1 page will jointly affect the customers' decision as whether to go to Product 2 page from Product 1 page. Other 2fi's in Table 8 can be interpreted similarly. By changing the level settings of the five factors,  $t_{12}$  can be as high as 0.2 or as low as 0.

We repeat the simulation for 10,000,000 times and record the conversions for each funnel separately. The calculated conversion rates are given in Table 9. Note that  $CF_{21}$  and  $CF_{23}$  both have no conversions in the simulated results.

The total conversion rate is assumed to be

$$CR_T = 2CR_1 + 2CR_3 + CR_{21} + CR_{2B} + CR_{21} + CR_{23}.$$

We model the conversion rate for each conversion funnel separately. Recall that the first step in modelling the conversion rate is to identify its related factors. For  $CR_1$ , it is the conversion rate for  $CF_1$ , which consists of three pages: the landing page, individual page and Product 1 page. Therefore, the six factors (A, B, C, D, G and H) related to these three pages are considered in modelling  $CR_1$ . Similarly, we can find the related factors for other conversion rates, all of which are listed in Table 10.

The model for each conversion rate is then built with respect to its related factors listed in Table 10. We consider the linear regression model with only the main effects and 2fi's and use the same method as in the toy example. The fitted model for each conversion rate is shown in Table 11.  $CR_{21}$  and  $CR_{23}$  are both zero because  $CF_{21}$  and  $CF_{23}$  have no conversions for all 64 versions of the conversion system in the simulated results.

In theory, Product 2 page can have four different versions, which would give us six more factors  $I'$ ,  $J'$ ,  $I''$ ,  $J''$ ,  $I'''$  and  $J'''$  in the optimisation procedure. This is because  $v_2$  appears in four different funnels, and the pages before  $v_2$  are different for each funnel. However, the factors related to  $v_2$  do not appear in any of the models in Table 11. Therefore, such consideration becomes unnecessary. Finally, by putting the functions

**Table 9.** Conversion rates for different funnels, simulated example.

Run	$CR_1$	$CR_3$	$CR_{21}$	$CR_{2B}$	$CR_{21}$	$CR_{23}$
1	0.011559	0.00073	0	0	0	0
2	0.011608	0	0	0	0	0
3	0.010784	0	0	0.002698	0	0
4	0.011201	0.00253	0	0	0	0
5	0.011635	0.002146	0	0	0	0
6	0.011559	0.003635	0	0	0	0
7	0.011191	0	0	0.001426	0	0
8	0.012039	0.001333	0	0	0	0
9	0	0	0	0	0	0
10	0	0.002268	0	0	0	0
11	0	0	0	0.001311	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0.002778	0	0
16	0	0	0	0	0	0
17	0	0.001971	0	0	0	0
18	0	0.030319	0	0	0	0
19	0	0	0	0.008195	0	0
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	0	0.010542	0	0	0	0
23	0	0.003757	0	0.003858	0	0
24	0	0.00412	0	0	0	0
25	0.059645	0	0	0	0	0
26	0.062159	0.006101	0	0	0	0
27	0.0621	0	0	0.003884	0	0
28	0.060896	0	0	0	0	0
29	0.061177	0	0	0	0	0
30	0.061716	0.018862	0	0	0	0
31	0.06134	0	0	0.008334	0	0
32	0.061483	0	0	0	0	0
33	0	0	0.004004	0	0	0
34	0	0.003807	0	0	0	0
35	0	0.003262	0.018645	0	0	0
36	0	0	0.003678	0	0	0
37	0	0.000699	0	0	0	0
38	0	0.012743	0	0	0	0
39	0	0.001638	0	0	0	0
40	0	0	0	0	0	0
41	0	0	0.010877	0	0	0
42	0	0.00712	0.011838	0	0	0
43	0	0	0	0	0	0
44	0	0	0	0	0	0
45	0	0	0	0	0	0
46	0	0.002403	0	0	0	0
47	0	0	0	0	0	0
48	0	0	0	0	0	0
49	0	0.007058	0.010615	0	0	0
50	0	0.012452	0	0	0	0
51	0	0	0.010327	0	0	0
52	0	0	0	0	0	0
53	0	0.002347	0	0	0	0
54	0	0	0	0	0	0
55	0	0.005054	0	0	0	0
56	0	0.005172	0	0	0	0
57	0	0	0.002712	0	0	0
58	0	0	0.002895	0	0	0
59	0	0	0.00266	0	0	0
60	0	0	0.002505	0	0	0
61	0	0	0	0	0	0
62	0	0.007343	0	0	0	0
63	0	0	0	0	0	0
64	0	0	0	0	0	0

**Table 10.** Conversion rates and related factors, simulated example.

$CR_1$	ABCDGH
$CR_3$	ABEFKL
$CR_{2I}$	ABCDIJ
$CR_{2B}$	ABEFIJ
$CR_{21}$	ABCDGHIJ
$CR_{23}$	ABEFIJKL

**Table 11.** Fitted models for each conversion rate, simulated example.

$$\begin{aligned}
 CR_1 &= 0.0091 + 0.0091A - 0.0062B - 0.0062C - 0.0091G - 0.0062AB - 0.0062AC + 0.0091AG \\
 CR_2 &= 0.0025 + 0.0017E - 0.0016F + 0.0013K - 0.0016EF - 0.0016FL \\
 CR_{2j} &= 0.0013 - 0.0013A + 0.0013D - 0.0013AD \\
 CR_{2B} &= 0.0005 + 0.0005A - 0.0005E + 0.0005F - 0.0005AE + 0.0005AF - 0.0005EF \\
 CR_{21} &= 0 \\
 CR_{23} &= 0
 \end{aligned}$$

in Table 11 together, we have the model for the total conversion rate to be

$$\begin{aligned}
 CR_T &= 0.025 + 0.0174A - 0.0124B \\
 &\quad - 0.0124C + 0.0013D + 0.0029E \\
 &\quad - 0.0027F - 0.0182G + 0.0026K \\
 &\quad - 0.0124AB - 0.0124AC - 0.0013AD \\
 &\quad - 0.0005AE + 0.0005AF + 0.0182AG \\
 &\quad - 0.0037EF - 0.0032FL.
 \end{aligned}$$

To maximise  $CR_T$ , we try all possible combinations of the nine factors involved in the above model. The maximal  $CR_T$  value is achieved by setting A, E, K and L to +, and B, C and F to -. The level settings of D and G do not affect the value of  $CR_T$  if the remaining seven factors are chosen as above. The other three factors that do not appear in the above model can be set based on other considerations.

## 6. Concluding remarks

In this work, we propose a new framework and approach to analyse a system of pages that relate to the conversion of interest based on the concept of conversion funnels. Directed graph is used to represent the system and identify all the conversion funnels. Fractional factorial design is used to conduct the experiment. An analysis strategy consists of modelling the conversion rate of each funnel separately and putting them together in the total conversion rate to do optimisation.

So far the analysis strategy is developed in the context of a specific example. But the underlying ideas are general. It is our plan to further develop this framework into a general methodology.

Though the general idea of this methodology may seem intuitive and straightforward as illustrated by the examples provided, its implementation can be challenging as the size of the web system increases. For example, suppose the number of webpages to be studied is now 20, which is quite typical for a real world web system, and suppose each page has only 2 factors, each with 2 levels, and that there are only 20 conversion funnels in this system. There are two immediate challenges facing the user: how to find a feasible design and how to optimise the overall conversion rate. In this case, the total number of factors need to be included in the design is 40 and the total number of factors need to be considered in the optimisation can easily reach hundreds due to the complex structure of the conversion funnels

and that some of the factors will need to have multiple versions as they would appear in different conversion funnels. For the design problem, it is not easy to find a 2-level fractional factorial design with 20 factors that is good according to the maximum resolution or minimum aberration criterion and whose run size is feasible for web system experimenters. Therefore, one might need to consider nonregular designs such as the Plackett-Burman designs or other orthogonal arrays. See the book by Wu and Hamada (2009) for details. In fact, since there are only limited numbers of interactions that may be important to the users, we can also use designs with relatively low resolution but has a few clear 2fi's to study the web system. For the second challenge, as the number of factors in the optimisation problem increases, exhaustive search might be infeasible so the user might want to try some other optimisation algorithms such as response surface or even integer programming. This problem becomes more daunting if the factors can have different numbers of levels because there is only a small dictionary of mixed-level orthogonal arrays with a large number of factors.

## Acknowledgments

This research is supported in part by ARO grant W911NF-17-1-0007. The authors are grateful to the Associate Editor and a referee for helpful comments.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

ARO [grant number W911NF-17-1-0007].

## Notes on contributors

*Heng Su* is a quantitative associate at Wells Fargo Bank.

*C. F. Jeff Wu* is the coca-cola chair in Engineering Statistics in Engineering Statistics at the School of Industrial and Systems Engineering, Georgia Institute of Technology.

## References

- Ash, T. (2009). *Landing page optimization: The definitive guide to testing and tuning for conversions*. Indianapolis, IN: Wiley.
- Donato, D., Laura, L., Leonardi, S., & Millozzi, S. (2004). Large scale properties of the webgraph. *The European*

*Physical Journal B - Condensed Matter and Complex Systems*, 38(2), 239–243.

Montgomery, D. C. (2012). *Design and analysis of experiments* (8th ed.). New York, NY: Wiley.

Ozolins, M. (2012). What is a conversion funnel. Retrieved 6 June, 2014, from <http://www.webics.com.au/blog/conversion-tracking/conversion-funnel/>

Qualaroo, (2014). *The beginner's guide to conversion rate optimization* (2nd ed.). Retrieved 6 June, 2014, from <https://qualaroo.com/beginners-guide-to-cro/user-experience-and-funnel-optimization/>

Wu, C. F. J., & Hamada, M. S. (2009). *Experiments: Planning, analysis, and optimization*. New York, NY: Wiley.