Working Paper, December 18, 2020, pp. XX–XX

Computer Resources for Implementing the Recipe Design for Weights and Scale Values in Multiattribute Judgment

Michael H. Birnbaum*

Abstract

In Multiattribute judgment, people form overall evaluations or decisions based on several cues or attributes. In order to measure the weights (importance) of the attributes and the scale values of the levels of the attributes, it is necessary to employ special designs. Factorial designs that includes all combinations of the levels of the attributes and representative designs that include combinations of the levels that occur in nature do not permit one to test between additive and averaging models, nor do they allow one to disentangle weights and scale values. Special experimental designs are required. The Recipe design is a formula for the case of three factors that guarantees that weights and scale values can be estimated in the averaging model and that the model can be tested. This paper presents three computer resources to assist researchers in learning about this design and model. The free online program, Recipe_wiz.htm, is a JavaScript powered Web page that makes Web forms for collecting Online data in a Recipe design. The program, Recipe_sim.htm, is a program that simulates data for the design according to a constant-weight averaging model. The Excel file, Recipe_fit.xlsx, is an Excel Workbook that uses the Solver to fit empirical (or simulated) data to the model in an example Recipe design. These resources, along with instructions and examples, are available at the following URL:
http://psych.fullerton.edu/mbirnbaum/recipe/

Keywords: Averaging Models, Conjoint Measurement, Functional Measurement, Importance of Variables, Information Integration, Multi-attribute utility, Recipe design, weights of attributes

1 Introduction

Most psychologists are familiar with analysis of variance (ANOVA) for a factorial design and with additive multiple linear regression for a "representative design" or "hybrid design" (Brunswik, 1956; Dharmi & Mumpower, 2018; Hammond, 1966; Postman & Tolman, 1959). Most know how these techniques have been used to estimate the effects of variables in an additive model. They know how these techniques can assess the additive model by testing the significance of interactions that violate the additive model.

However, I suspect that many students of psychology are less familiar with the use of averaging models to analyze multiattribute judgments (Anderson, 1974, 1981). Further, I think many do not realize that special experimental designs and quantitative techniques are required in order to measure the weights (importance) of variables, and scale values (the subjective values of levels of the variables) in an averaging model.

Studies based on simple factorial or representative designs, analyzed via ANOVA or regression are simply not capable of testing between adding and averaging models; furthermore, these techniques do not permit unambiguous estimation of scale values and weights.

Some people are not aware that statistical models such as additive multiple regression may or may not be descriptive of empirical results, and that their application for multiattribute judgments may be inappropriate. Some students are confused by the use of the terms "regression weights" and "beta weights", which seem to indicate these might be indices of the importance of variables, but these numbers are not the same as weights in the averaging model. I think that the pun of using the term "weights" for the multiplicative coefficients in regression has created some confusion. Regression weights are not really identifiable distinct from the variance of the levels used in the analysis, so even if the models were descriptive, these "weights" would not represent "importance" of variables (Birnbaum & Stegner, 1981).

Indeed, when multiple regression weights have been estimated from data and compared with judged "importance" of variables, there has not been much agreement between these indices (e.g., Nisbett & Wilson, 1977). The lack of correlation between these arbitrary indices even led some people to think that people cannot judge what is important to them (Nihm, 1984).

In order to teach students about experimental design, av-

I thank Bonny Quan and Trinity Tay for useful discussions of the issues in this paper.

*Department of Psychology, California State University, Fullerton. Email: mbirnbaum@fullerton.edu.

eraging models, and estimation of weights and scale values, the Recipe design and an associated computer program in FORTRAN were created in 1976 as a modification of the program used in Birnbaum (1976). The idea was to provide a "recipe" for designing a study that would guarantee that a student could test between additive and averaging models, and also would be able to estimate weights and scale values in the constant-weight averaging model.

A computer program, RECIPE, was written for "functional measurement", or estimation of the weights and scale values according to the constant-weight averaging model under the assumption that the response scale is a linear function of subjective value (Birnbaum, 1976). Similar but slightly different techniques and experimental designs were presented by Norman (1976).

The Recipe design and the RECIPE program have been used in a number of doctoral dissertations, Masters' theses and student projects over the years. Some of these studies were reviewed in Birnbaum and Stegner (1981), Birnbaum and Mellers (1983) and Stevenson, Busemeyer, and Naylor (1990).

Empirical tests of the constant-weight averaging model led to evidence of violations of the relative-weight averaging models with constant weights, used in Anderson's (1981) "functional measurement" approach. Evidence of systematic violations led to configural weight theories (Birnbaum, 1974, 1982), which led to other methods for studying phenomena that violate the constant weight models (Birnbaum & Jou, 1990; Birnbaum & Zimmermann, 1998; Birnbaum, 2008).

Configural weight averaging models violate some of the implications of the constant-weight "information integration" models of Anderson (1974, 1981), including the implication of no interaction in the two-way and three-way designs. However, they agree in the implications that test between adding and averaging models and they give very similar estimates of the importance of factors.

Therefore, even though there is empirical evidence against constant-weight averaging models in many situations (Birnbaum, 1982), this class of models remains a useful and important framework and null hypothesis for the study of multiattribute evaluative judgments.

This paper provides updated resources useful to those who wish to learn this approach to research design and analysis, to provide tools that can provide stepping stones to more advanced work that build upon this foundation. There are three main programs, together with instructions and examples.

These three programs are called, respectively, Recipe_Wiz.htm (creates Web forms to collect data online), Recipe_sim.htm (simulates data in a Recipe design according to the constant-weight averaging model with error), and Recipe_fit.xlsx (an Excel workbook that illustrates how to use the Solver in Excel to estimate the weights and scale values in the constant-weight averaging model and to create the diagnostic graphs that allow one to compare data with the best-fit "predictions" of the parameters. The program, Recipe_Wiz.htm, is similar to Birnbaum's (2000) FacorWiz.htm program, for which many examples and applications are described in Birnbaum (2001).

These resources are freely available Online at the following URL:
http://psych.fullerton.edu/mbirnbaum/recipe/

1.1 Recipe Design

The Recipe design is based on three factors, designated A, B, and C, with n_A, n_B and n_C levels. It consists of the union of the 3-way factorial design of A by B by C, denoted ABC, combined with each 2-way factorial design (with one piece of information left out), denoted AB, AC, and BC, combined with each piece of information presented alone: A, B, and C. There are a total of $(n_A+1)(n_B+1)(n_C+1)-1$ experimental trials ("cells") in the Recipe design.

The following hypothetical example study can be used to illustrate a Recipe design: Judgments of intentions to accept a new COVID-19 vaccine. At the time this article was written (Fall, 2020), a deadly viral disease, COVID-19, had killed 240,000 Americans – about 4% of those confirmed to have contracted it – and many of those who recovered suffered lingering health problems. No treatments or vaccines had yet been approved but a great deal of misinformation about treatments and vaccines was circulating, and there was distrust in all sources of information, because of political interference in the health agencies and disinformation campaigns presented in the presidential election of 2020. The question was: If a new vaccine were developed and approved, would people decide to take it? There was a concern that not enough people might decide to accept vaccination to suppress the pandemic.

In this vaccination example, let A = Price, B = Risk (of side effects of the vaccine), and C = Effectiveness (of the vaccine to prevent infection if exposed to the virus). It is important to distinguish an attribute (a cue or factor, like Price) from its levels. Levels of the attribute of Price might include $10, $500, etc.

If there are 3 levels of price, 4 levels of Risk, and 5 levels of Effectiveness, there would be 119 cells (experimental trials) in the design, consisting of 3 trials for the three levels of Price alone, 4 trials of Risk alone, 5 trials of Effectiveness alone, 3 by 4 = 12 trials of Price combined with risk, etc.

These 119 trials will be intermixed and presented in random order (following a suitable warm-up of representative trials). The dependent variable in this example would be a rating of "how likely would you be to take the new vaccine?"

1.2 Averaging Model

The relative-weight averaging model with constant weights (Anderson, 1974; Birnbaum, 1976; Norman, 1976) can be written for a three attribute situation as follows:

$$ABC_{ijk} = \frac{w_0 s_0 + w_A a_i + w_B b_j + w_C c_k}{w_0 + w_A + w_B + w_C} \quad (1)$$

where ABC_{ijk} is the theoretical response in the case where all three attributes are presented, with levels i, j, and k, respectively, which have scale values of a_i, b_j, and c_k, respectively. The weights (importance) of factors A, B, and C are w_A, w_B, and w_C, respectively. The initial impression has a weight of w_0 and a value of s_0. In theory, the value of the initial impression, s_0, represents the value of the impression in the absence of information, and its weight, w_0, represents how resistant this impression is to new information. The initial impression is analogous to the concept of a prior in Bayesian reasoning, but it functions differently in the averaging model from what is implied by the standard application of Bayesian model (Birnbaum & Mellers, 1983).[1]

A key assumption of the averaging model is that if an attribute is not presented, its weight is zero. For example, if Factor C is not presented, the theoretical impressions (based only on A and B) are given as follows:

$$AB_{ij} = \frac{w_0 s_0 + w_A a_i + w_B b_j}{w_0 + w_A + w_B} \quad (2)$$

where AB_{ij} is the theoretical response based on A and B (where C has not been presented), with the levels of i and j, respectively. Note that Equation 2 follows from Equation 1 by simply setting w_C to zero.

Similarly, for judgments based on A alone (leaving out both B and C), $w_B = w_C = 0$, leading to the following:

$$A_i = \frac{w_0 s_0 + w_A a_i}{w_0 + w_A} \quad (3)$$

The marginal means for any factor are linearly related to the scale values for that factor in any design in which that factor appears, so the judgments of A when A is presented alone, the marginal means for A when A is presented with B, or the marginal means for A when presented with C or with both B and C, for example, should all be linearly related to each other For example, in the AB design, the marginal means for A, averaged over levels of B are given as follows,

$$\overline{AB}_{i\bullet} = a_i \frac{w_A}{w_0 + w_A + w_B} + \frac{w_0 s_0 + w_B \overline{b}}{w_0 + w_A + w_B} \quad (4)$$

where $\overline{AB}_{i\bullet}$, are the marginal means for A in the AB design, and $\overline{b}$ is the mean scale value of Factor B. From these expressions, we can in principle solve for the scale values. In principle, because the Recipe design affords many estimates of the same parameters, it means that the design imposes multiple constraints on the same parameters, which allows one to test the implications of the model.

Note that the effect of changing the level of A (the coefficient of a_i in Equations 1 – 4 is inversely related to the total weight of other factors presented with A. This inverse relationship leads to the concept of the Zen of Weights, described in the next section.

1.3 Zen of Weights

In additive models such as ANOVA or multiple regression, in factorial, representative, or hybrid designs, people measure the main effects of a variable and compare effects. Sometimes people say that a variable that has greater effects is more "important".

In the averaging models, however, the effect of a variable is not to be confused with its weight. One cannot not examine the effects of A to determine its weight. Instead, one examines the effects of A to determine the weights of B and C. Because this approach seems paradoxical at first (especially to those familiar with ANOVA or multiple regression), this concept is called the "Zen of weights."

To understand the Zen of weights, it is useful to introduce some notation. We will refer to the *effect of A* as the difference in response as the factor A is manipulated from A_1 to A_m. These equations hold for any two levels of A, but it might be helpful to think of levels 1 and m as those that create the lowest and highest responses for A. The indices, i, j, and k will be used for the levels of A, B, and C, respectively, and a bullet ($\bullet$) will be used to denote that responses have been averaged over levels of a factor.

When A is presented alone, the effect of A is defined as follows:

$$\Delta A = A_m - A_1 \quad (5)$$

The effect of A in the AB design, denoted $\Delta A(B)$, is defined as the difference in marginal means for A in the AB design for A; that is,

$$\Delta A(B) = \overline{AB}_{m\bullet} - \overline{AB}_{1\bullet} \quad (6)$$

Where $\overline{AB}_{i\bullet}$ denotes the marginal mean in the AB design for level i of A, averaged over the levels of B.

The effect of A in the AC design is denoted $\Delta A(C)$, and is defined as follows:

$$\Delta A(C) = \overline{AC}_{m\bullet} - \overline{AC}_{1\bullet} \quad (7)$$

Finally, the effect of A in the ABC factorial design, denoted $\Delta A(BC)$, is given by,

$$\Delta A(BC) = \overline{ABC}_{m\bullet\bullet} - \overline{ABC}_{1\bullet\bullet} \quad (8)$$

According to the additive model, all of these effects are assumed to be equal; however, according to the constant-weight averaging model (Equation 1), these effects of A

[1]This equation and the following treatment implicitly assumes that the judgment function, J, which maps subjective impressions to overt responses, is linear. In either linear case or when J is only monotonic, the scale values are determined in theory to an interval scale, and the weights can be multiplied by any factor, k, without altering the order of the data.

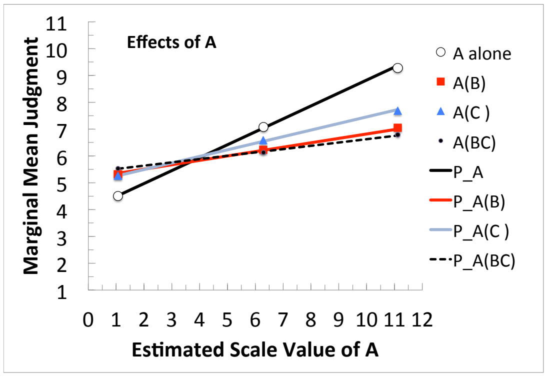

FIGURE 1: Marginal Means for Factor A in A alone, AB, AC, and ABC designs. Markers represent means from simulated data and lines show best-fit predictions of the theory. From the effects of A it can be seen that in this case, B has a greater weight than C. One can also draw similar figures for the effects of B and C.

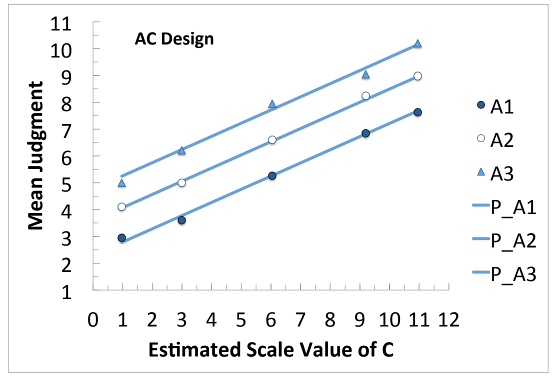

FIGURE 2: Mean judgments in the AC design, plotted as a function of estimated scale values of C, with a separate curve for each level of A; markers show mean judgments and lines show best-fit predictions of the model.

are inversely related to the total weight of the information presented. That is,

$$\Delta A = \Delta a \frac{w_A}{w_0 + w_A} \tag{9}$$

$$\Delta A(B) = \Delta a \frac{w_A}{w_0 + w_A + w_B} \tag{10}$$

$$\Delta A(C) = \Delta a \frac{w_A}{w_0 + w_A + w_C} \tag{11}$$

$$\Delta A(BC) = \Delta a \frac{w_A}{w_0 + w_A + w_B + w_C} \tag{12}$$

where the range of scale values, $\Delta a = a_m - a_1$, is the same in all expressions, but the relative weights are different. According to this model, the $\Delta A(BC)$ will be the smallest, and ΔA will be greatest. The key implication of the model is as follows:

$$\Delta A(B) < \Delta A(C) \iff w_B > w_C \tag{13}$$

In words, to compare the weights of B and C, we examine the effects of A. if the weight of B is greater than the weight of C, then the effect of A will be less when B is presented with it than when C is presented with A. The more important a variable is, the less the effects of other variables when it is included.

Figure 1 plots the predicted marginal means in Equations 5, 6, 7, and 8, according to the model of Equation 1, as a function of the scale values for A; i.e., a_i. The markers represent simulated data that contain error, and are discussed in the section on Recipe_sim.xlsx below. Note that the curve for A alone has the steepest slope and the curve for A(BC) has the least slope. In this case, the slope for A(B) is less than that for A(C), so here we see $w_B > w_C$.

These expressions are similar for the effects of B and Cm and we can draw figures like Figure 1 for each of these factors. From the effects of B, B(A) and B(C), we can compare the relative weights of A and C. Finally, we can compare the weights of A and B by examining the effects of C(A) and C(B).

A very basic implication of the model that can be tested in the Recipe design is as follows: If the $\Delta A(B) < \Delta A(C)$ and $\Delta B(C) < \Delta B(A)$, it follows that $\Delta C(B) < \Delta C(A)$. It should be clear that there is redundancy in the design, so the weights can not only be estimated from these slopes, but the model can be tested by this design.

Among other implications, the relative weight averaging model with constant weights implies that there should be no two-way interaction between any factors in the AB, AC, or BC designs, and no two-way or three-way interactions in the ABC design (Anderson, 1974, 1981). Figure 2 shows the AC design, with lines showing predicted judgments as a function of the estimated scale values of C, with a separate curve for each level of A. The curves are parallel.

If overt responses are monotonically but not linearly related to subjective impressions, interactions could occur, even if the constant-weight model were correct (Birnbaum, 1974, 1982). But even with a nonlinear judgment function, the model implies that in any two-way design, curves cannot cross, they cannot trend in opposite directions, they must satisfy double cancellation, and they must satisfy joint independence in the ABC design (aka, "branch independence" in some contexts). For more on the ordinal implications, see Krantz, Luce, Suppes, and Tversky (1971); for connections between interactions and these ordinal properties, see Birnbaum & Zimmermann (1998).

Empirically, many studies have observed interactions that have been interpreted as evidence of configural weighting (Birnbaum, 1974; 2008; Birnbaum & Stegner, 1979, 1981; Birnbaum & Zimmermann, 1998). Although these violations of the constant-weight averaging model refute the simple constant-weight averaging model (that predicts parallel lines in Figure 2), they appear compatible with a configural weight averaging model that has the same implications regarding the relative effects of variables in the recipe design (the slopes in Figure 1).

In sum, if we use the Recipe design, we can estimate the weights and scale values and test the averaging model. But before we can fit data to the model, we must collect data. The next section describes a program that can be used to construct experiments that can collect data via the WWW.

2 Recipe_Wiz.htm

The Web page, Recipe_Wiz.htm, contains a JavaScript program that allows the user to create Web pages containing a form to collect data in the Recipe design. It is similar to FactorWiz (Birnbaum, 2000). All of the open-source code is contained in the same page, so it can easily be modified or expanded for specialized purposes.

To use the program for the first time, visit the page and note that default values have already been introduced in the fields for the example study of decisions whether or not to accept a new COVID-19 vaccine. The three variables in the default study are A = Price, B = Risk (of harmful side-effects), and C = Effectiveness (of the vaccine to prevent COVID).

To set up a new experiment, one should select the text in the "Experiment Name" box and type in a name that will be displayed for the participant; this name can contain spaces. For the example study, this title might be, "Will you take the new COVID-19 vaccine?" Then use your computer's tab key to tab to the "short name" field, and type in a name such as you might use for the file name of the study; this name should contain no spaces and it will be saved as the first variable in the data file. For the example, it might be "vaccine_01".

Pushing the computer's tab key again will select the name of the name of Factor A. This name will appear in the created Web form and it can contain HTML. The default given is, PRICE: ; the tags and are HTML for bold text. Pressing tab each time, or clicking and selecting the fields, you can type in the information for Factors B and C.

By pushing the tab key or by selecting a field in the table, one can enter the levels of the factors. The program allows one to enter material that is to appear before each trial ("Preliminary Material"), after each trial ("After Material"), and as separators between the factors. This material can be HTML, including links to images or other Web-supported media. The default values are
 in each of these cases;
 is the HTML tag for line return.

Recipe_Wiz.htm is designed to provide a response scale consisting of a row of radio buttons; the user can specify the number of buttons and the labels to appear at the endpoints of the scale. In the default for the example, there are 11 buttons labeled from "very very unlikely" to "very very likely" (to try the vaccine).

When all the fields in the table have been filled in, the user can press the button labeled "Make the form" and the HTML will appear in the large (textarea) box. Pressing the button "Display Form" will open a new window and display the questionnaire that has been created. It will contain a four-trial warm-up and the trials for the study (in random order), as well as a list of standard demographic questions such as age, gender, education, etc. Pressing "Make the form" again will create the same form, except with the experimental trials in a different random order.

Pressing the "Save" button displays a pop-up message that informs the user that the text in the box should be saved in a text editor (not a word processor), and it selects all the text. One can then use CTRL-C and CTRL-V to copy and paste this text into a text editor, where it can be saved and edited. The file name should have no spaces in the name and should have an extension of .htm or .html. For example, the filename might be "vaccine_01.htm", corresponding to the short name. This file could then be uploaded to the WWW, where participants can complete the questionnaire.

When the participant presses the submit button, the data will be sent to a PERL script and saved to a file at the following URL:
http://ati-birnbaum.netfirms.com/data/data.txt

One need only change the ACTION of the FORM tag to direct the data to another server. Instructions for running your own server and for installing a generic PERL script to save data to your own server are given in Reips and Birnbaum (2011).[2]

Even though the trials are in random order, the data are organized by the generic.pl script in the proper order for Recipe analysis, starting in the 12th column. The first columns list the short experiment name, date, time, IP address, and other background information. Starting in the 12th column the responses are ordered by designs; the designs are ordered: A, B, C, AB, AC, BC, and ABC. Within designs, they are ordered in factorial order with the last-listed factor index moving the fastest. For example, for a 3 by 4, AB design the order is $AB_{11}, AB_{12}, AB_{13}, AB_{14}, AB_{21}, \ldots, AB_{n_A n_B}$, where n_A and n_B are the number of levels of A and B, respectively.

In the HTML appears "(put your instructions here)", where

[2]It is advised to save data from any real experiment in a secure location, rather than on the Web, as was done here. This exception has been set up to allow people to test the system and view sample data, rather than for real data collection.

the researcher can paste the instructions. It is recommended to edit the warm-up trials to include the worst and best cases and a representative sample of perhaps $7-11$ warm-up trials, to allow the participant to adapt to the stimuli and the response scale. An example study including instructions and such a warm-up is linked from the Website where the resources are listed:
http://psych.fullerton.edu/mbirnbaum/recipe/

3 Recipe_sim.htm

The Web page, Recipe_sim.htm, contains a JavaScript program that simulates data for a Recipe design according to the relative-weight averaging model. This program is designed to be compatible with Recipe_Wiz.htm, and it produces data that can be analyzed in the same way as empirical data generated by the forms created by Recipe_Wiz.htm.

The purpose of the simulation program is to enable researchers and students who want to understand the model to be able to adjust parameters to explore predictions of the model for data. The simulation program also makes explicit exactly how the model functions.

The program contains defaults for the same hypothetical example that make it easy to use for the first time. One can again enter a title and a short name, which might include words such as "Sim" to remind some future viewer of the output that these are simulated data. In the first line, one can specify how many simulated "subjects" should be created by the program.

One can then enter names for Factors A, B, and C, the weights of the factors (w_A, w_B, w_C), the numbers of levels of each factor(n_A, n_B, n_C), the scale values of the factors (a_i, b_j, c_k), and the weight and scale value of the initial impression (w_0, s_0).

It is assumed that responses in this task are a linear function of the subjective impressions given by the averaging model. To allow for predictions to different rating scales, the linear coefficients of this judgment function can be changed from the defaults of an identity function.

Computer generated, "random" error can be added to each judgment. The program implements errors that are normally distributed with a mean of zero, and the standard deviation of the normal distribution can be specified by the user. Finally, the user can also specify the precision with which the simulated responses should be rounded (e.g., 7.62, 7.6, or 8).

When these values have been set, pushing the "Simulate Many Cases" button will perform the calculations and enter the results in the box. Pushing the "Save" button selects all the data in the window, and a pop-up reminds the reader to copy and paste these (using CTRL-C and CTRL-V) into some program such as Excel or into a text file, where they could be saved as a comma separated values (.csv) file. The predictions start in the 12th position and are in the same arrangement as those created by the forms made via Recipe_Wiz.htm.

In the first columns, the words "date", "time", "IP", etc. are placed in the file where actual dates and times, etc. would have been placed in an empirical study from a form created by Recipe_Wiz.htm.

Pushing the "Label Data" will record in the data box the information that the user has used in the simulation, allowing the user to copy and paste this information to keep a record of such information as the names and number of levels of the factors used in the simulation. Because only some of the information is saved, a user doing a number of related simulations might be advised to take a screen shot of the window in order to keep a complete record of the parameters specified.

4 Recipe_fit.xlsx

The Excel Workbook, Recipe_fit.xlsx, is designed to illustrate how to fit data (either empirical data from participants in an experiment or simulated data) to the model. This workbook is set up to analyze data from a 3 by 4 by 5 design, and is compatible with the default experiment in Recipe_Wiz.htm and the default settings in Recipe_sim.htm. A user who can generalize from this example and who understands Excel can modify this worksheet for other designs. Some lessons on using Excel to analyze data including the Solver can be found in Birnbaum (2001); there are many other free resources on the Web for learning Excel.

The program accepts raw data in the form generated by either the experimental forms generated by Recipe_Wiz.htm or Recipe_sim.htm. It finds mean judgments, organizes them by design, finds the appropriate marginal means, and via the Solver, it finds the best-fit (least squares) parameters for a set of data for the relative-weight averaging model. It constructs graphs of the effects of A alone, A(B), A(C), A(BC), B alone, B(A), B(C), B(AC), C alone, C(A), C(B), and C(AB). It also plots the factorial graphs of AB, AC, and BC factorial designs, and of the ABC design.

The Solver does not always appear in an Excel installation, and it may need to be installed, usually from Excel's "Add-ins". Even without the Solver in Excel, the Workbook can be used to find marginal means and construct predictions including graphs of those predictions and of the data. With the Solver, the program can be used to estimate the best-fit parameters and corresponding predictions.

5 Discussion

These programs should prove useful for researchers, instructors, and students who wish to apply, teach, and learn about how to test theories of multiattribute judgment, to measure "importance" of variables, and the properties of averaging models as theoretical representations of these tasks.

In the past, much confusion was created by researchers who mistakenly believed that regression "weights" or ANOVA effects were estimates of the "importance" of variables, and a literature developed that people do not know what is "important" to them (Nihm, 1984; Nisbett & Wilson, 1977). Much of the confusion arose because weights in additive models are not really identifiable and additive models are not descriptive (Anderson, 1981).

Research has shown that averaging models and the Zen of weights are far more accurate as descriptive models of human behavior and weights in these models seem better correlated with judgments of the importance of variables (Birnbaum & Mellers, 1983; Birnbaum & Stegner, 1981).

The Recipe design is not the only design in which weights and scale values can be estimated, and the models can be tested. Once a person has mastered the concepts, the person may devise more efficient designs to can accomplish specific goals. An example in which there were five factors, each with five levels is provided in Birnbaum and Stegner (1981). In that study, people predicted the IQ of adopted children based on factors such as the biological parents' IQs, adopting parents' IQs, and the socioeconomic status of the adopting environment. To include all RECIPE combinations of 5 levels of 5 factors, as in RECIPE, would have required $6^5 = 7775$ trials. A smaller design with 200 trials was designed that sufficed for the purposes of that study.

In the Recipe design the inter-correlations of variables are zero, because the design consists of the union of factorial designs. If we wish to test the effects of the correlations among factors (e.g., to test Brunswik's (1956) ideas), one can embed the Recipe design among contextual trials to manipulate the correlations. A design in which the context (including correlation) is systematically manipulated is known as "systextual" design (Birnbaum, 1975, 2007; Stevenson, et al., 1990). One can then test theories of correlations among variables to generalize to any environment (Birnbaum, 2007).

5.1 Configurally weighted averaging models

Evidence in certain studies revealed systematic deviations from the relative weight averaging models with constant weights. These phenomena have been fit by means of configural weighting models, in which the relatively higher or lower valued information may have greater weight (Birnbaum, 2008; Birnbaum & Stegner, 1979; Birnbaum & Zimmermann, 1998).

The findings of interactions by Birnbaum, Parducci, and Gifford (1981), Birnbaum (1972; 1973; 1974), Birnbaum and Veit (1974), and others led to disputes with Norman H. Anderson, who had published data that seemed to satisfy this prediction. These controversies are reviewed in Section F of Birnbaum (1982). An issue of debate was whether or not the interactions might be attributed to a nonlinear function between subjective impressions and the overt response. However, converging evidence from new experimental tasks and constraints showed that the interactions are "real" (Birnbaum, 1974; Birnbaum & Jou, 1990; Birnbaum & Zimmermann, 1998).

Birnbaum (1972, 1973, 1974) and Birnbaum and Jou (1990) found that judgments of the likeableness of a person described by a set of adjectives or the morality of a person who has done a number of good or bad deeds do not conform to the parallelism prediction, as in Figure 2. Instead, when one of the adjectives or deeds is bad, the other information has less effect. Such an interaction can be described by a configural weight model in which the worst personality trait or moral deed has greater weight than the best trait or deed. By analogy, I would predict that judgments of likelihood to take a new vaccine would be low if the price is high, if the risk is high, or if the effectiveness is low.

When the dependent variable is an evaluation from bad to good and the judge is in the point of view of evaluating whether to accept (e.g., to buy something, to marry someone, to accept an applicant for a job, etc.), the violations have been interpreted as a person placing greater configural weight on the lower, more unfavorable information (Birnbaum, 1974).

For example, people appear to place greater weight on an adjective such as "malicious" than on the adjective "sincere" when evaluating a person who is "malicious and sincere"; people tend to place a greater weight on an estimate of \$500 for the value of a used item than on an estimate provided by an equally expert source who says it is worth \$1500. Risk aversion in gambles has also been explained by configural weighting (Birnbaum, 2008): the reason people prefer \$45 over a 50-50 gamble to win either \$1 or \$100 is that they place greater weight on the lower outcome.

But there are cases where people seem to put greater weight on the higher-values information. That tends to occur when the judge is in the "seller's point of view". When asked what is the least one would accept to sell an item like a used car or a gamble, people tend to place greater weight on the higher valued information (Birnbaum, 2018; Birnbaum & Stegner, 1979; Birnbaum & Sutton, 1992; Birnbaum, Coffey, Mellers, & Weiss, 1992; Birnbaum, Yeary, Luce, & Zhao, 2016).

5.2 Concluding Comments

The old RECIPE program in FORTRAN proved very effective in the 1970s and 1980s for teaching students how to design and analyze a study in which weights can be separated from scale values and to test models of multi-attribute judgment. It is hoped that these updated resources should make it again convenient to accomplish these research and educational goals.

References

Anderson, N. H. (1974). Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & D. Suppes (Eds.), *Contemporary developments in mathematical psychology*, vol. 2. San Francisco: W. H. Freeman.

Anderson, N. H. (1981). *Foundation of information integration theory*. New York: Academic Press.

Birnbaum, M. H. (1972). Morality judgments: Tests of an averaging model. *Journal of Experimental Psychology, 93*(1), 35-42.

Birnbaum, M. H. (1973). Morality judgment: Test of an averaging model with differential weights. *Journal of Experimental Psychology, 99*(3), 395-399.

Birnbaum, M. H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology Monograph, 102*, 543-561.

Birnbaum, M. H. (1975). Expectancy and judgment. In F. Restle, R. Shiffrin, N. J. Castellan, H. Lindman, & D. Pisoni (Eds.), *Cognitive Theory, Vol. 1* (pp. 107-118). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Birnbaum, M. H. (1976). Intuitive numerical prediction. *American Journal of Psychology, 89*, 417-429.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp 401-485). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science, 10,* 399-407.

Birnbaum, M. H. (2000). SurveyWiz and FactorWiz: JavaScript Web pages that make HTML forms for research on the Internet. *Behavior Research Methods, Instruments, and Computers, 32*, 339-346.

Birnbaum, M. H. (2001). *Introduction to Behavioral Research on the Internet*. Upper Saddle River, NJ: Prentice Hall.

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology, 55*, 803-832.

Birnbaum, M. H. (2007). Designing online experiments. In A. Joinson, K. McKenna, T. Postmes, & U.-D. Reips (Eds.), *Oxford Handbook of Internet Psychology* (pp. 391-403). Oxford, UK: Oxford University Press.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review, 115*, 463-501.

Birnbaum, M. H. (2018). Empirical evaluation of third-generation prospect theory. *Theory and Decision, 84*(1), 11-27.

Birnbaum, M. H., & Bahra, J. P. (2012a). Separating response variability from structural inconsistency to test models of risky decision making, *Judgment and Decision Making, 7*, 402-426.

Birnbaum, M. H., & Jou, J. W. (1990). A theory of comparative response times and "difference" judgments. *Cognitive Psychology, 22*(2), 184-210.

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology, 45*, 792-804.

Birnbaum, M. H., Parducci, A., & Gifford, R. K. (1971). Contextual effects in information integration. *Journal of Experimental Psychology, 88*(2), 158-170.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37*, 48-74.

Birnbaum, M. H., & Stegner, S. E. (1981). Measuring the importance of cues in judgment for individuals: Subjective theories of IQ as a function of heredity and environment. *Journal of Experimental Social Psychology, 17*, 159-182.

Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes, 52*(2), 183-215.

Birnbaum, M. H., & Veit, C. T. (1974). Scale-free tests of an averaging model for the size-weight illusion. *Perception & Psychophysics, 16*(2), 276-282.

Birnbaum, M. H., Wong, R., & Wong, L. (1976). Combining information from sources that vary in credibility. *Memory & Cognition, 4*(3), 330-336.

Birnbaum, M. H., Yeary, S., Luce, R. D., & Zhao, L. (2016). Empirical evaluation of four models for buying and selling prices of gambles. *Journal of Mathematical Psychology, 75*, 183-193.

Birnbaum, M. H., & Zimmermann, J. M. (1998). Buying and selling prices of investments: Configural weight model of interactions predicts violations of joint independence. *Organizational Behavior and Human Decision Processes, 74*(2), 145-187.

Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley, CA: University of California Press.

Dhami, M. K., & Mumpower, J. L. (2018). Kenneth R. Hammond's contributions to the study of judgment and decision making. *Judgment and Decision Making, 13*(1), 1–22.

Hammond, K. R. (1966). Probabilistic functionalism: Egon Brunswik's integration of the history, theory and method of psychology. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik*. New York: Holt, Rinehart, & Winston.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement*. New York: Academic Press.

Nihm, S. D. (1984). Self-reports on mental processes: A response to Birnbaum and Stegner. *Bulletin of the Psychonomic Society, 22*, 426-427.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259.

Norman, K. L. (1976). A solution for weights and scale values in functional measurement. *Psychological Review, 83*, 80-84. DOI: 10.1037/0033-295X.83.1.80

Postman, L. & Tolman, E. C. (1959): Brunswik's probabilistic functionalism. In: S. Koch (Ed.), *Psychology: A Study of a Science, Study 1: Conceptual and Systematic, Vol. 1: Sensory, Perceptual, and Physiological Formulations* (pp. 502-564). New York: McGraw-Hill.

Reips, U.-D., & Birnbaum, M. H. (2011). Behavioral research and data collection via the Internet. In R. W. Proctor & K.-P. L. Vu (Eds.), *Handbook of human factors in Web design* (2nd Edition) (pp. 563-585). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Riskey, D. R., & Birnbaum, M. H. (1974). Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology, 103*(1), 171-173.

Stevenson, M. K., Busemeyer, J. R., & Naylor, J. C. (1990). Judgment and decision-making theory. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (p. 283–374). Consulting Psychologists Press.