

# Using Excel to estimate parameters from observed data: An example from source memory data

C. S. DODSON, W. PRINZMETAL, and A. P. SHIMAMURA  
*University of California, Berkeley, California*

We present an application, using Excel, that can solve best-fitting parameters for multinomial models. Multinomial modeling has become increasingly popular and can be used in a variety of domains, such as memory, perception, and other domains in which processes are assumed to be dissociable. We offer an application that can be used for a variety of psychological models and can be used on both PC and Macintosh platforms. We illustrate the use of our program by analyzing data from a source memory experiment.

Models of psychological phenomena often include the identification of underlying factors or unobservable processes. In such models, it is assumed that a dependent variable, such as reaction time or proportion correct, can be represented by a number of different parameters, such as processing speed, amount of information encoded, decision criterion, and guessing biases. In memory research, mathematical models have been used to contrast or dissociate memory processes, such as familiarity versus recollection (Jacoby, 1991), recognition search versus decision criterion (Atkinson & Juola, 1974), and item versus source memory (Batchelder & Riefer, 1990). For example, in recognition memory experiments involving source memory, one can identify and dissociate parameters associated with item memory (e.g., determining whether a test item was old or new) from parameters associated with source identification (e.g., whether an item was presented by a male or female).

Multinomial models have provided one method of estimating parameters associated with psychological phenomena, including source memory (see Riefer & Batchelder, 1988, for a fuller discussion). From behavioral data, such as proportion correct on memory tests, it is possible to estimate parameters associated with performance, using maximum likelihood estimation or other *best-fit* procedures. Microsoft Excel, a commonly used spreadsheet application, includes an iterative function called *Solver* that can be used to solve many parameter estimation problems. In this report, we demonstrate the use of Excel's *Solver* function to derive multinomial parameters. This procedure has several advantages: (1) the program is widely available; (2) it can be used on both Macintosh and PC computers; (3) it is easy to implement; and (4) it employs an efficient

algorithm. We will illustrate the use of *Solver* in the multinomial analysis of data from a source memory experiment.

## THE SOURCE MONITORING EXPERIMENT

In a typical source monitoring experiment, subjects learn information from different sources and later are required to remember the source of the test items (see Johnson, Hashtroudi, & Lindsay, 1993, for a review). For instance, the subjects could hear words spoken by a man (i.e., source A) and a woman (i.e., source B). On a subsequent source test, words from both sources are mixed with new distracter words. For each test word, the subjects must make a three-alternative forced-choice: Was the test word spoken earlier by the man or the woman, or is it a new word? The data from the source test can be summarized in a table, such as Table 1, in which the row headings correspond to the actual source of the test item and the column headings correspond to the subject's response to the test item. The summary data for responses to the words from each source, such as the male, can be represented by the following probabilities: (1) the source identification rate is represented by the probability of correctly stating that a word spoken by the male came from that source (i.e.,  $P\{\text{"Male"} \mid \text{Male}\}$ ); (2) the source confusion rate is represented by the probability of incorrectly stating that a word spoken by the male was spoken by the female (i.e.,  $P\{\text{"Female"} \mid \text{Male}\}$ ); and (3) the miss rate is represented by the probability of incorrectly stating that a word spoken by the male was a new word (i.e.,  $P\{\text{"New"} \mid \text{Male}\}$ ).

Batchelder, Riefer, and colleagues assume that the responses on a source test depend on the cognitive state of the subject with respect to each test word. These cognitive states and the probabilities of making various responses are represented by different parameters within the multinomial model that is schematically displayed in Figure 1 (adapted from Batchelder & Riefer, 1990). Figure 1 displays the processes that are associated with re-

---

We thank Peter Dixon for originally suggesting the *Solver* function for data fitting and David Riefer for his valuable suggestions on the manuscript. Correspondence concerning this article should be addressed to C. Dodson, Department of Psychology, University of California, Berkeley, CA 94720 (e-mail: chadd@violet.berkeley.edu).

**Table 1**  
**Summary of Response Outcomes From a**  
**Three-Alternative Forced-Choice Source Test**

True Source	Response		
	Male	Female	New
Male	<b>P("Male"   Male)</b>	P("Female"   Male)	P("New"   Male)
Female	P("Male"   Female)	<b>P("Female"   Female)</b>	P("New"   Female)
New	P("Male"   New)	P("Female"   New)	<b>P("New"   New)</b>

Note—Correct responses are in bold.

sponding to both studied items (i.e., items from sources A and B) and new items. Current multinomial models of source monitoring (see, e.g., Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996; Riefer, Hu, & Batchelder, 1994) assume that the cognitive states represent two different memory processes: (1) *Item detection* refers to the memorial information that allows studied words to be distinguished from new words on the test, and (2) *source identification* refers to the memorial information that identifies the source of studied words.

As can be seen in Figure 1, subjects detect a studied word as *old* with probability  $D_i$  where  $i$  indicates source  $i$ . Then, they either remember the source of the word with

probability  $d_i$  and correctly identify it, or they fail to remember any source information about the word and guess its source with probability  $a$  for source A and  $1 - a$  for source B. However, if a studied word is not initially detected as *old*, then, with probability  $b$  (see Figure 1), the subject *guesses* the word is an old word whose source has been forgotten. With probability  $g$ , the subject guesses that the item is from source A, and, with probability  $1 - g$ , the subject guesses the item is from source B. Figure 1 also displays the processes that are involved in responding to new items. Subjects identify new words as *new* with probability  $1 - b$ . However, with probability  $b$ , subjects incorrectly respond *old* to a new item (i.e., a false alarm).

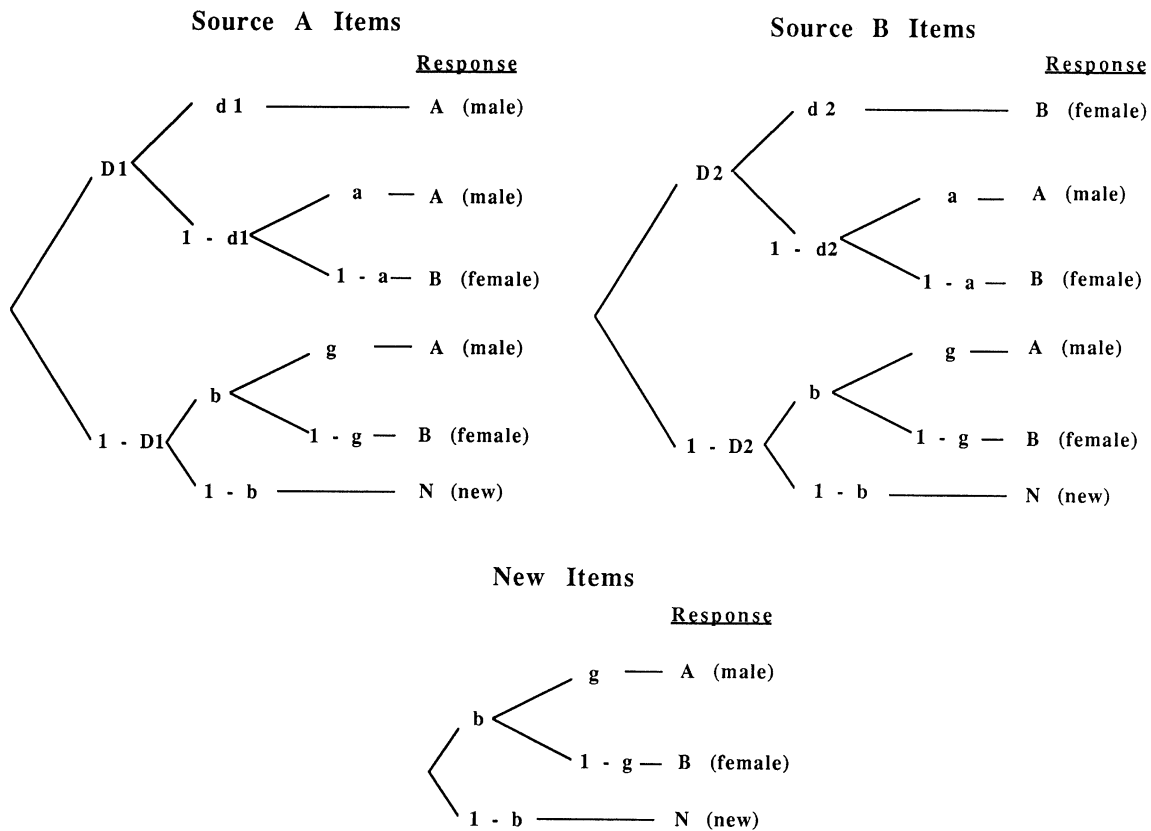


Figure 1. Tree diagrams for the multinomial model, with separate trees for items said by the male source and for new items.  $D_1$  = probability of detecting male items as old;  $d_1$  = probability of identifying the source of male items;  $a$  = probability of guessing that a detected item was spoken by the male;  $b$  = probability of guessing an item is old;  $g$  = probability of guessing that an undetected item was spoken by the male. From "Multinomial Processing Models of Source Monitoring," by W. H. Batchelder and D. M. Riefer, 1990, *Psychological Review*, 97, p. 551. Copyright 1990 by the American Psychological Association. Adapted with permission.

In summary, the different memory states and response biases are:

(1) *Item detection*: In Figure 1, the parameter  $D_i$  is the probability that an item from source  $i$  is correctly detected as old, regardless of the accuracy of source identification.

(2) *Source identification*: The parameter  $d_i$  is the probability of correctly identifying the source of an item from source  $i$ , given that it has been correctly detected as old.

(3) *Response Biases*: There are three different types of response biases in these models.

(3a) The parameter  $b$  is the response bias (i.e., probability) to respond *old* to undetected old items and to new items.

(3b) When a subject has guessed that an item is *old* (i.e., the  $b$  probability), the parameter  $g$  is the probability that the subject guesses that the item is from source A.

(3c) When a subject correctly detects an *old* item as old but does not remember the source of the item, the parameter  $a$  is the probability that this subject says that the item is from source A.

Thus, for each response category in Table 1, we can derive a prediction from the tree diagram illustrated in Figure 1. For example, given items originally spoken by the male (source A items), the probability of responding *male* is simply the sum of all of the paths leading to this response:

$$p(\text{"male"} | \text{male}) = (D.1 * d.1) + (D.1 * [1 - d.1] * a) + ([1 - D.1] * b * g). \tag{1}$$

According to the model, there are three different paths that will result in a subject's responding *male* to items spoken by the male. First, the test item may feel *old* to the subject and he or she recollects that the male earlier spoke the item [i.e.,  $(D.1) * (d.1)$ ]. Second, the test item feels *old*, and the subject cannot recollect who spoke the item but correctly guesses that the male spoke it [i.e.,  $(D.1) * (1 - d.1) * (a)$ ]. Third, the subject guesses that the test item is old and also correctly guesses that the male spoke it [i.e.,  $(1 - D.1) * (b) * (g)$ ]. For each response category (Table 1), there is a similar formula that can be derived to predict the observed proportion of responses. Finally, it is necessary to find values for the parameters (e.g.,  $D.1$ ,  $d.1$ , etc.) that provide the best fit of the data. The best-fitting parameter values can be found with the Solver function in Excel, as is illustrated below.

**SOLVING THE MULTINOMIAL FUNCTIONS**

To illustrate the method of computing parameter values for the multinomial model, we use data from an experiment by Dodson, Holland, and Shimamura (1997). In this study, the subjects heard words spoken by two different voices. The subjects in the low-similarity condition heard a male and a female (referred to as Eric and Julie) say words, whereas the subjects in the high-similarity condition heard two different males (referred to as Eric and Anthony) say words. Finally, the subjects completed a

three-alternative forced-choice source test. For each test word, the subjects decided whether the word was new or old, and, if old, who said it. Table 2A presents the results from the low-similarity (Eric/Julie condition), and Table 2B presents the results from the high-similarity (Eric/Anthony) condition. We will first consider only the low-similarity condition (Table 2A).

We have developed an Excel spreadsheet that analyzes data, like the kind in Table 2A, with a multinomial model. We describe the spreadsheet here (it also can be downloaded at <http://socrates.berkeley.edu/~teachp/>).<sup>1</sup> Our spreadsheets in Figures 2A and 2B only work for experiments containing data from three different sources, such as the data in Table 2A. However, as we will discuss later, it is easy to modify the spreadsheet in order to analyze data from more than three sources (see, e.g., Dodson, Holland, & Shimamura, 1998; Riefer et al., 1994). We will first define the contents in each of the columns, and then we will describe a procedure for computing parameter values.

The eight different columns in the spreadsheet are defined as follows:

- (A) The parameter names.
- (B) The parameter values. These cells have been given names, as described below.
- (C) The name of the item's source.
- (D) The name of the probability of a particular response.
- (E) The observed frequency of a response.
- (F) The actual probability of a response.
- (G) The multinomial model's equations for predicting the probability of a response.
- (H) The formula for calculating  $G^2$ , which measures the goodness of fit between the observed probability of a response and the model's predicted probability of a response.

Using Excel's Define Name function (in the Insert menu for Version 5), we have named cells B5 to B11 to correspond to the parameter names. For example, cell B7 has been named "ld.1." The value in this cell corresponds to the parameter value for  $d.1$  (in Figure 1). Similarly, cell B9 has been named "a" and its value corresponds to the

**Table 2A**  
**Response Frequencies for the Low-Similarity Voices Condition**

Item	Response		
	Eric	Julie	New
Eric	<b>612</b>	151	77
Julie	123	<b>643</b>	74
New	19	18	<b>383</b>

Note—Correct responses are in bold.

**Table 2B**  
**Response Frequencies for the High-Similarity Voices Condition**

Item	Response		
	Eric	Anthony	New
Eric	<b>521</b>	210	109
Anthony	172	<b>566</b>	102
New	28	32	<b>360</b>

Note—Correct responses are in bold.

A

	A	B	C	D	E	F	G	H
1								
2	Parameters							
3								
4	Names	Values	Source	Response	Obsvtns	Actual	Predicted Probability	G ^ 2
5	D.1	.500	Eric	p("Eric")	612	.729	= (D.1*ld.1)+(D.1*(1-ld.1)*a)+((1-D.1)*b*g)	= (2*E5)*(LN(F5/G5))
6	D.2	.500		p("Julie")	151	.180	= (D.1*(1-ld.1)*(1-a)) + ((1-D.1)*b*(1-g))	= (2*E6)*(LN(F6/G6))
7	ld.1	.500		p("New")	77	.092	= (1-D.1) * (1-b)	= (2*E7)*(LN(F7/G7))
8	ld.2	.500	Julie	p("Eric")	123	.146	= (D.2*(1-ld.2)*a) + ((1-D.2)*b*g)	= (2*E8)*(LN(F8/G8))
9	a	.500		p("Julie")	643	.765	= (D.2*ld.2)+(D.2*(1-ld.2)*(1-a))+((1-D.2)*b*(1-g))	= (2*E9)*(LN(F9/G9))
10	g	.500		p("New")	74	.088	= (1-D.2) * (1-b)	= (2*E10)*(LN(F10/G10))
11	b	.500	New	p("Eric")	19	.045	= b * g	= (2*E11)*(LN(F11/G11))
12				p("Julie")	18	.043	= b * (1-g)	= (2*E12)*(LN(F12/G12))
13				p("New")	383	.912	= 1 - b	= (2*E13)*(LN(F13/G13))
14								
15							G^2 =	sum(H5:H13)
16								
17								
18								
19								

B

	A	B	C	D	E	F	G	H
1								
2	Parameters							
3								
4	Names	Values	Source	Response	Obsvtns	Actual	Predicted Probability	G ^ 2
5	D.1	.500	Eric	p("Eric")	612	.729	.500	460.809
6	D.2	.500		p("Julie")	151	.180	.250	-99.608
7	ld.1	.500		p("New")	77	.092	.250	-154.509
8	ld.2	.500	Julie	p("Eric")	123	.146	.250	-131.591
9	a	.500		p("Julie")	643	.765	.500	547.695
10	g	.500		p("New")	74	.088	.250	-154.370
11	b	.500	New	p("Eric")	19	.045	.250	-64.962
12				p("Julie")	18	.043	.250	-63.489
13				p("New")	383	.912	.500	460.310
14								
15							G ^ 2 =	800.285
16								
17								
18								
19								

Figure 2. Sample spreadsheet for computing parameter values for a multinomial model of an experiment with three sources.

parameter a. Columns A, C, and D in Figures 2A and 2B play no functional role in the analysis, but they make the output more readable. Also, the spreadsheets in Figures 2A and 2B display the contents of the spreadsheet before applying Excel's Solver program. Importantly, starting values need to be entered in column B before running the Solver program. We have used values of .500, but these values are arbitrary and any set of numbers ranging between 0 and 1 will work.

As can be seen in Figure 2A, the Predicted Probability column contains the model's equations for determining a subject's response to a test item. Figure 2A displays precisely what must be entered into each cell. Because Excel converts the equations in columns G and H into numerical values (on the basis of the parameter values in column B), Figure 2B shows the values of these equations in each cell. In short, each cell in the Predicted Probability column contains equations that represent the possible ways in which a particular response can occur in terms of the tree diagram, shown in Figure 1. For example, as seen in Figure 2A the first cell contains the following equation:

$$= (D.1 * ld.1) + (D.1 * (1 - ld.1) * a) + ((1 - D.1) * b * g). \tag{2}$$

This formula is simply a rewrite of Equation 1, using the names we have given the cells. There is a formula for each of the nine response categories in this experiment.

The task now is simply to find parameter values (Figure 2, column B) so that the predicted probabilities (column G) match the observed proportions (column F) as closely as possible. What is required is a measure of goodness of fit. We use the maximum likelihood estimator  $G^2$ , as recommended by Riefer and Batchelder (1988), but other measures of goodness of fit could be used (e.g.,  $SSE$ ,  $R^2$ ,  $\chi^2$ ).

The goodness-of-fit measure,  $G^2$ , is defined as follows:

$$\sum (2 * ObsFreq_i) * \ln \left( \frac{ObsProp_i}{ObsPred_i} \right). \tag{3}$$

For each response category  $i$ , ObsFreq is the observed frequency, ObsProp is the observed proportion, and Obs-

Pred is the predicted proportion. Cells H5 to H13 are the terms of the summation, and cell H15 is the  $G^2$  value. A small difference between the observed and predicted probabilities is, of course, a good fit and results in a small  $G^2$ . Excel uses its Solver program to estimate best-fitting parameter values by minimizing the value of  $G^2$  in row 15, column H.

Before running Solver, it is necessary to make some additional decisions. The overall seven-parameter model, presented in Figure 1, is not identifiable because there are more parameters (seven) than there are degrees of freedom in this data set (six). Therefore, it is necessary to select a submodel of the overall model with six or fewer parameters. Batchelder and Riefer (1990) present seven possible submodels of the full seven-parameter model that were created by imposing one or more of the following parameter restrictions:  $D.1 = D.2$ ,  $d.1 = d.2$ , and  $a = g$ .

When choosing an appropriate model, our general strategy is to select a model with the fewest number of parameters that adequately fit the data. Batchelder and Riefer (1990) recommend the strategy of first determining whether or not  $D.1$  can be equated with  $D.2$ . In other words, is the old–new recognition rate equivalent for items from the two sources? A model-free test of this strategy is to compare the probability of responding *new* to items from the two sources. For instance, in Table 2A, in the low-similarity condition, the frequency of responding *new* to Eric items is almost equal to this frequency for Julie items (i.e., 77 vs. 74, respectively). Therefore, we selected a model of the low-similarity data that set  $D.2$  equal to  $D.1$ . To set  $D.2$  equal to  $D.1$ , we inserted “= D.1” in cell B6, thus equating  $D.2$  and  $D.1$ . Figure 2B presents a four-parameter model of our low-similarity data. In this model,  $D.2 = D.1$ ,  $d.2 = d.1$ , and  $g = a$ . That is, on the spreadsheet in Figure 2B, we have set the following three sets of values equal to each other: First,  $D.2$  in cell B6 equals the value of  $D.1$  in cell B5; second, the value of  $d.2$  in cell B8 equals the value of  $d.1$  in cell B7; and third, the

value of  $g$  in cell B10 equals the value of  $a$  in cell B9. When entering starting values in column B, be careful not to disrupt the cell entries for setting the parameters equal to each other. Figure 2B displays what the spreadsheet looks like before running Solver.

### USING EXCEL'S SOLVER PROGRAM

To test the goodness of fit of the model we use Excel's Solver function from the Tool menu (Excel Version 5) or the Formula menu (Excel Version 4). If Solver does not appear in this menu, it can be found by clicking the Add-in option of this menu. When using this function, it is necessary to specify three variables in the Solver dialog box. Figure 3 illustrates the Solver dialog box. There are three crucial components of this box.

(1) The Set Target Cell option: This option refers to the cell in the worksheet whose value is minimized (or maximized, or made to reach a specified value). For the model presented in Figure 2B, the target cell is the value in Row 15, column H, because this is the model's  $G^2$  value that must be *minimized*.

(2) The Changing Cells option: This option refers to those cells in the worksheet whose values are adjusted until a solution is found. In Figure 2B, some of the parameters in column B are designated to change in order to minimize the value of  $G^2$  in cell H15. For the four-parameter model in Figure 2B, the parameters that are specified to change are  $D.1$  (row 5, column B),  $d.1$  (row 7, column B),  $a$  (row 9, column B), and  $b$  (row 11, column B). The appropriate parameter names should be entered in this option.

(3) The Subject to the Constraints option: By using the Add, Change, and Delete buttons of the dialog box, it is possible to set constraints on the values of the cells in the worksheet. For instance, it is necessary to set the constraint that the parameter values in column B can only vary between the values of 0 and 1. Because we want our param-

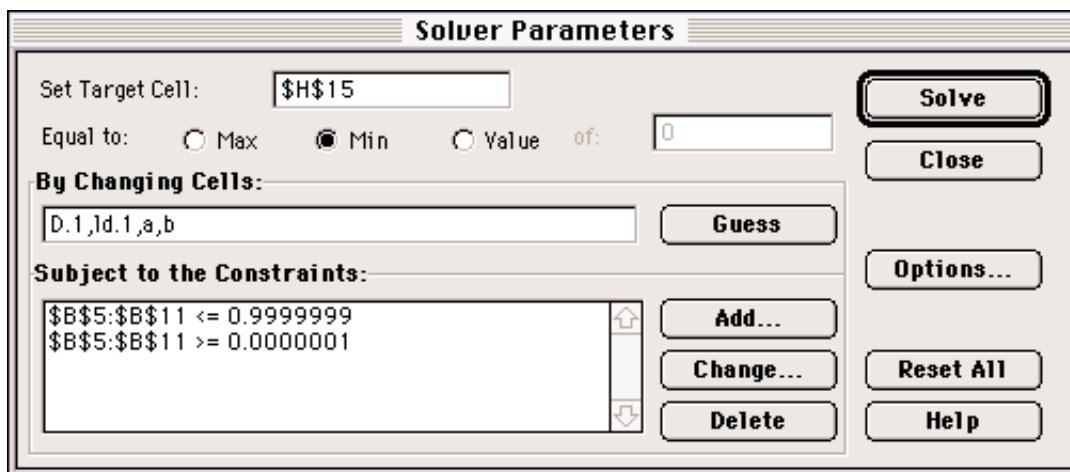


Figure 3. The Solver dialog box.

	A	B	C	D	E	F	G	H
1								
2								
3	Parameters				Probability			
4	Names	Values	Source	Response	Observations	Actual	Predicted	G ^ 2
5	D.1	.901	Eric	p("Eric")	612	.729	.732	-5.912
6	D.2	.901		p("Julie")	151	.180	.178	2.941
7	ld.1	.648		p("New")	77	.092	.090	3.029
8	ld.2	.648	Julie	p("Eric")	123	.146	.148	-3.147
9	a	.454		p("Julie")	643	.765	.762	6.183
10	g	.454		p("New")	74	.088	.090	-2.971
11	b	.088	New	p("Eric")	19	.045	.040	4.640
12				p("Julie")	18	.043	.048	-4.123
13				p("New")	383	.912	.912	.000
14								
15							G^2 =	.641
16								
17								
18								
19								
20								

Figure 4. Four-parameter model of the low-similarity data.

eters to be probabilities and  $G^2$  is not defined when the predicted probability is 0, we restrict the parameters to vary from “.0000001” to “.9999999.”

After inputting these three components and clicking on the Solve button, the spreadsheet should look like the spreadsheet in Figure 4. The parameter values in column B are the values that yielded the lowest  $G^2$  and, consequently, the best fit of the data to the particular model. We found that the four-parameter model in Figure 4 fit the data in the low-similarity condition and contained the fewest parameters.

One might question whether this fit is a good fit. With the data in Table 2A, the resulting value of  $G^2$  in cell H15 is 0.641. The value of  $G^2$  is interpreted as a chi-square value because  $G^2$  is distributed asymptotically as a chi-square distribution. When the value of  $G^2$  is greater than a critical value, it is necessary to reject the model. With two degrees of freedom (i.e., six degrees of freedom minus four free parameters) and an alpha level of .05, the critical value on a chi-square table is 5.99. Thus, we cannot reject the four-parameter model of our low-similarity data because 0.641 is less than 5.99 (i.e., [ $G^2(2) = 0.641, p > .05$ ]). We also used a four-parameter model for the high-similarity data, and it also fit the data [ $G^2(2) = 0.342, p > .05$ ]. Rather than describing the fit of a model to a set

of data with a goodness-of-fit measure such as  $G^2$  or percent of variance accounted for, it is more satisfying to contrast two models and determine whether one provides a significantly better account of the data than does the other. We will demonstrate this later. However, when examining the fit of one model, it is very useful to look for systematic deviations between the actual and predicted probabilities (i.e., Figure 4, columns F and G).

Solver uses the method of gradient descent (i.e., a quasi-Newton method) to find optimal parameters. Occasionally, Solver will encounter a local minima and report a solution that is not the best fit, given the data and model. To prevent this problem, Solver should be run several times with different starting parameter values in column B (not all .500, as in Figure 2). Our experience with simple multinomial models for source memory is that local minima are rarely encountered. However, with more complex problems, we have occasionally encountered local minima. If one suspects such problems, one can start Solver with many different starting values and choose the fit with the lowest  $G^2$ . An Excel macro that iteratively runs Solver with random parameter starting values is illustrated on the Web site <http://socrates.berkeley.edu/~teachp/>

Hu (1991) has created a program that uses the EM algorithm instead of the method of gradient descent and is

Table 3  
A Comparison of the Hu (1991) and Solver Programs in the Analysis of Six Different Data Sets

Data Set	Hu's Program					Solver Program				
	Parameter Estimates				Goodness of Fit	Parameter Estimates				Goodness of Fit
	D	d	g	b	$G^2(2)$	D	d	g	b	$G^2(2)$
Low similarity	.901	.648	.454	.088	0.641	.901	.648	.454	.088	0.641
High similarity	.853	.491	.450	.143	0.342	.853	.491	.450	.143	0.342
1	.267	.951	.477	.462	1.544	.267	.951	.477	.462	1.544
2	.673	.876	.556	.313	0.817	.673	.876	.556	.313	0.817
3	.748	.640	.508	.454	0.634	.748	.640	.508	.454	0.634
4	.863	.647	.546	.390	0.0003	.863	.647	.546	.390	0.0003

Note—Data sets 1, 2, 3, and 4 are taken from Tables 5 and 7 from Batchelder and Riefer (1990).

specifically designed for analyzing data with multinomial models (see Hu & Batchelder, 1994, for a discussion of the EM algorithm). Because Hu's program is the standard for analyzing data with multinomial models, in Table 3 we present a comparison of the results of Hu's program and the Solver program. For the six different data sets, we obtained identical parameter estimates and values of  $G^2$  with both programs. Thus, although the two programs employ different algorithms, they produce the same results.

There are advantages and disadvantages to both Hu's program and using Solver. Excel is widely available and runs on a variety of platforms, whereas Hu's program is only available on PCs. There are a number of nice features of Hu's program, such as its ability to derive confidence intervals for parameter estimates, test models for identifiability, check if the model equations are correct, and so forth. However, Hu's program is limited to binary tree branches. Solver, in contrast, can handle very complicated tree structures (see, e.g., Dodson et al., 1998) and can combine multinomial approaches with those based on continuous distributions (see, e.g., Ashby, Prinzmetal, Ivry, & Maddox, 1996). Confidence intervals for parameter estimates can be derived with Solver, using a boot-

strapping procedure, as is illustrated by a spreadsheet on the Web site.

**TESTING QUESTIONS OF INTEREST**

The procedure for testing experimental questions is similar to the procedure for testing the goodness of fit of a model. In this example, we ask whether source similarity affects the source identification rate, as measured by the value of  $d$ . Basically, the idea behind testing the statistical significance of differences in parameters across experimental conditions involves comparing the goodness of fit of two different models. In one model, both parameters can vary independently of each other, and, in the other model, both parameters are set equal to each other. If the model that allows both parameters to vary independently fits the data significantly better than does the model in which both parameters are set equal, it is possible to conclude that the values of the parameters are significantly different. Conversely, if the fits of both models are not significantly different, the parameters are not significantly different. This general procedure can be used to compare any nested models—that is, any situation in which one model is a special case of another.

	A	B	C	D	E	F	G	H
1	Low Similarity Condition							
2								
3	Parameters				Probability			
4	Names	Values	Source	Response	Observations	Actual	Predicted	G ^ 2
5	D1.low	.901	Eric	p("Eric")	612	.729	.732	-5.912
6	D2.low	.901		p("Julie")	151	.180	.178	2.941
7	ld1.low	.648		p("New")	77	.092	.090	3.029
8	ld2.low	.648	Julie	p("Eric")	123	.146	.148	-3.147
9	a.low	.454		p("Julie")	643	.765	.762	6.183
10	g.low	.454		p("New")	74	.088	.090	-2.971
11	b.low	.088	New	p("Eric")	19	.045	.040	4.640
12				p("Julie")	18	.043	.048	-4.123
13				p("New")	383	.912	.912	.000
14								
15	G^2 =							.641
16								
17								
18								
19	Total G^2							.983
20								
21								
22	High Similarity Condition							
23								
24	Parameters				Probability			
25	Names	Values	Source	Response	Observations	Actual	Predicted	G ^ 2
26	D1.hi	.853	Eric	p("Eric")	521	.620	.624	-6.415
27	D2.hi	.853		p("Anthony")	210	.250	.250	-.564
28	ld1.hi	.491		p("New")	109	.130	.126	7.114
29	ld2.hi	.491	Anthony	p("Eric")	172	.205	.205	.095
30	a.hi	.450		p("Anthony")	566	.674	.670	6.927
31	g.hi	.450		p("New")	102	.121	.126	-6.883
32	b.hi	.143	New	p("Eric")	28	.067	.064	2.054
33				p("Anthony")	32	.076	.079	-1.986
34				p("New")	360	.857	.857	.000
35								
36	G^2 =							.342

Figure 5. Four-parameter model of the low-similarity data and a four-parameter model of the high-similarity data.

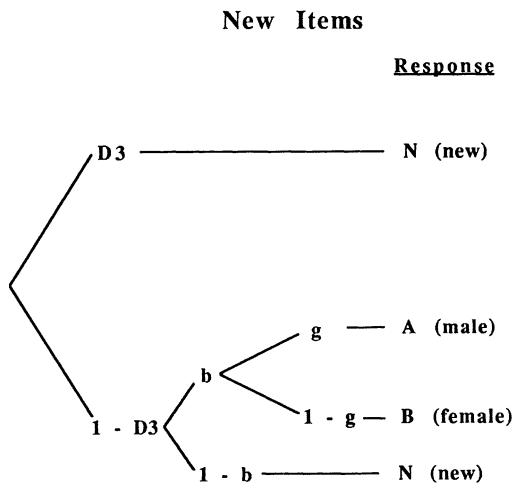


Figure 6. Tree diagram for the responses to *new* items in a two high-threshold multinomial model.  $D3$  = probability of detecting a *new* item as *new*;  $b$  = probability of guessing an item is old;  $g$  = probability of guessing that an undetected item was spoken by the male.

To investigate the source identification rate (i.e., the  $d$  parameter) in the low- and high-similarity conditions, we constructed a spreadsheet containing a four-parameter model of the low-similarity data and a four-parameter model of the high-similarity data, as is shown in Figure 5. Numerically,  $d$  is higher in the low-similarity condition (.65) than in the high-similarity condition (.49). But are these two values significantly different? The procedure for determining whether or not these values are significantly different involves the following three steps.

First, we combine both data sets (high and low source similarity) and find the best fit for an eight-parameter model to all the data combined (18 response categories, instead of 9). In Figure 5, the parameters in this model are labeled  $D.1.low$ ,  $ld.1.low$ ,  $a.low$ ,  $b.low$ ,  $D.1.hi$ ,  $ld.1.hi$ ,  $a.hi$ , and  $b.hi$ . It is necessary to use Excel's Define Name function to rename these parameters. As in the previous example, we reduced the total number of parameters by setting  $D.2.low = D.1.low$ ,  $ld.2.low = ld.1.low$ ,  $g.low = a.low$ ,  $D.2.hi = D.1.hi$ ,  $ld.2.hi = ld.1.hi$ ,  $g.hi = a.hi$ . Our  $G^2$  for the entire data set is in cell F20 (i.e., 0.983); this is the sum of the terms H5 to H13 (i.e., 0.641) and H26 to H 34 (i.e., 0.342). The parameter estimates and  $G^2$  value in Figure 5 are the values that appear after running Solver and finding the best-fitting model. To run Solver on this spreadsheet, open the Solver dialog box, make the target cell F20 (i.e.,  $G^2$ ), and specify the following eight cells to change:  $D.1.low$ ,  $ld.1.low$ ,  $a.low$ ,  $b.low$ ,  $D.1.hi$ ,  $ld.1.hi$ ,  $a.hi$ , and  $b.hi$  (the parameters). After running Solver, we note that  $G^2 = 0.983$  and the source discrimination parameters for the low- and high-similarity conditions are .648 and .491. These are, of course, reasonable values, because source discrimination should be more difficult if the sources of information are similar.

The important part of this first step is to note the  $G^2$  value from the best-fitting model (i.e., 0.983).

The second step is to evaluate the following seven-parameter model. The seven-parameter model is identical to the eight-parameter model, except that we set  $ld.1.hi = ld.1.low$ . That is, we are evaluating the fit of a model that assumes that these parameter values (.648 and .491) are actually the same. After setting the value of cell B28 to be  $ld.1.low$ , run Solver. In the By Changing Cells section of the Solver dialog box, it is necessary to change the following seven parameters:  $D.1.low$ ,  $ld.1.low$ ,  $a.low$ ,  $b.low$ ,  $D.1.hi$ ,  $a.hi$ , and  $b.hi$  (we have removed  $ld.1.hi$  to give us seven parameters). After running Solver, we find that the  $G^2$  of the seven-parameter model is 26.79. Of course, the eight-parameter model ( $G^2 = 0.983$ ) provides a better fit than the seven-parameter model ( $G^2 = 26.79$ ).

Finally, to determine whether the seven-parameter model is significantly worse than the eight-parameter model, subtract the two  $G^2$  values from the two different models (i.e.,  $26.79 - 0.983$ ), yielding a difference of 25.807. The difference can be interpreted as a  $\chi^2$  with 1  $df$ . Note that the degree of freedom is simply the difference in the number of parameters between the two models (i.e., eight parameters minus seven parameters). Since  $G^2(1)$  is greater than the critical value of 3.84 (critical value of  $\chi^2$  with 1  $df$ ), we can reject the seven-parameter model in favor of the eight-parameter model. Another way of stating this is that source discriminability in the low-similarity condition ( $dl.low$ ) is significantly worse than source discriminability in the high-similarity condition [ $dl.hi$ ;  $G^2(1) = 25.807$ ,  $p < .01$ ].

Note that the method for comparing models with differences in the value of  $G^2$  applies only to nested models—that is, models in which one model is a special case of a more general model. To test models that are not nested and have different numbers of parameters, other techniques must be used (see Ashby, 1992; Sakamoto, Ishiguro, & Kitagawa, 1986).

## MODIFYING THE MULTINOMIAL MODEL

Recently, Bayen et al. (1996) proposed a two high-threshold (2HT) multinomial model, and have shown that this model provides a more accurate analysis of the data than does the one high-threshold (1HT) model introduced by Batchelder and Riefer (1990). The only difference between these two models is in the treatment of new items. The tree diagram for the new items in the 2HT model is illustrated in Figure 6. The 2HT model proposes that, just as old items can be detected as *old* (i.e., with  $D.1$ ,  $D.2$ ), new items can be detected as *new* (i.e., with  $D.3$ ; see Bayen et al. for further details). Thus, there is an additional parameter,  $D3$  in the 2HT model (see Figure 6). Often,  $D3$  is equated with the  $D.1$  and  $D.2$ . To modify the 1HT model, shown in Figure 2A, into the 2HT model (Figure 6), it is necessary to change the equations in the last three rows of the Predicted Probability column. We



leave this as an exercise for the reader. However, for the four-parameter 2HT model of the low-similarity data, we obtained the following parameter values:  $D.1 = D.2 = D.3 = .82$ ;  $ld.1 = ld.2 = .71$ ;  $a = g = .45$ ;  $b = .50$ . Note that, on the basis of goodness of fit (i.e.,  $G^2$ ) alone, it is impossible to distinguish between the 1HT and 2HT models, because both models fit the data equally well. However, Bayen et al. provide evidence for favoring the 2HT model over the 1HT model.

Another recently proposed multinomial model by Dodson et al. (1998) measures memory for relatively exact and partial source information. In these experiments, there were four sources of information (items spoken by "Anthony," "Eric," "Julie," and "Lucy") in addition to new items. In this study, source information was not measured in an all-or-none manner—that is, the model contained parameters for measuring both voice and gender information. For example, subjects could remember that the item was spoken by a male but not remember whether it was spoken by Anthony or Eric. To modify the spreadsheet in Figure 2A, the additional parameter names and values, as well as the source names and responses, must be added to the appropriate columns. However, the major changes will be to the Predicted Probability column. See Figure 2 in Dodson et al. to determine the appropriate equations for this column.

The application of multinomial models, solved with Excel, is of course not limited to memory research. These source memory models trace their lineage to threshold models developed in studies of perception and used to correct for guessing biases (see, e.g., Link, 1982). One application from visual perception that is conceptually similar to the source monitoring experiment is the study of illusory conjunctions (see, e.g., Prinzmetal, 1995). In a typical illusory conjunction experiment, subjects are briefly presented with colored letters, such as a blue T and a red F. Veridical perception requires the correct combination of letters and colors. Errors that result from incorrectly combining a color and a letter are called illusory conjunctions (e.g., reporting a red T). Whereas source monitoring errors in memory reflect the incorrect combination of a remembered item and its source, illusory conjunction errors reflect the incorrect combination of visual features in perception. Using Excel and methods nearly identical to those described here, Prinzmetal and colleagues have modeled results from a variety of illusory conjunction experiments (see Ashby et al., 1996; Prinzmetal, Henderson, & Ivry, 1995). In general, multinomial models can be used whenever responses can be placed in nominal categories and the hypothesized psychological processes are stochastic. One model derived by Ashby et al. (1996) is of particular interest because it combines formulas derived from tree diagrams, as in Figure 1, with a continuous model akin to signal detection theory. Because Excel can calculate probabilities from several statistical distributions (normal, beta, gamma, poisson, etc.), such hybrid models are easy to manage.

## CONCLUSION

Although we have focused on an example from memory research, Excel's Solver program can be used to analyze *any* and *all* multinomial models.<sup>2</sup> The Excel program is an important tool for solving multinomial equations, because of its ease of use and its widespread availability for both Macintosh and PC computers. The Excel function Solver is a convenient tool for estimating parameters in multinomial models. However, its application is not limited to multinomial models. Solver can be used by any model in which a method of gradient descent can be used to optimize the goodness of fit between predicted and observed data. Many standard statistical procedures are examples of such problems. For example, in linear regression, model parameters (intercept and slope) are selected to minimize the sum-of-squares errors. With a convenient method of finding optimal parameters, the theorist is not limited to standard models, such as linear regression, but can derive models that fit complex experiments.

## REFERENCES

- ASHBY, F. G. (1992). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 31-34). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G., PRINZMETAL, W., IVRY, R., & MADDOX, T. (1996). A formal theory of illusory conjunctions. *Psychological Review*, **103**, 165-192.
- ATKINSON, R. C., & JUOLA, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1, pp. 243-293). San Francisco: Freeman.
- BATCHELDER, W. H., & RIEFER, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, **97**, 548-564.
- BAYEN, U. J., MURNANE, K., & ERDFELDER, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 197-215.
- DODSON, C. S., HOLLAND, P. W., & SHIMAMURA, A. P. (1997). *Speaker similarity and memory*. Unpublished raw data.
- DODSON, C. S., HOLLAND, P. W., & SHIMAMURA, A. P. (1998). The recollection of specific and partial source information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1-15.
- HU, X. (1991). *Statistical inference program for multinomial binary tree models* (Version 1.0) [Computer Program]. Irvine: University of California.
- HU, X., & BATCHELDER, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, **59**, 21-47.
- JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.
- JOHNSON, M. K., HASHTRUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- LINK, S. (1982). Correcting response measures for guessing and partial information. *Psychological Bulletin*, **92**, 469-486.
- PRINZMETAL, W. (1995). Visual feature integration in a world of objects. *Current Directions in Psychological Science*, **4**, 90-94.
- PRINZMETAL, W., HENDERSON, D., & IVRY, R. (1995). Loosening the constraints on illusory conjunctions: The role of exposure duration and attention. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1362-1375.
- RIEFER, D. M., & BATCHELDER, W. H. (1988). Multinomial modeling

- and the measurement of cognitive processes. *Psychological Review*, **95**, 318-339.
- RIEFER, D. M., HU, X., & BATCHELDER, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 680-693.
- SAKAMOTO, Y., ISHIGURO, M., & KITAGAWA, G. (1986). *Akaike information criterion statistics*. Boston: Reidel.

#### NOTES

1. The spreadsheets were created with Excel 4.0, so they can be read by any version (PC or Mac) that is 4.0 or later. The download files con-

tain four spreadsheets that illustrate the following problems: (1) the low-similarity condition discussed in the text; (2) a comparison of the low- and high-similarity conditions discussed in the text; (3) an example illustrating a macro that iteratively selects random starting values in situations in which local minima are suspected; and (4) an example illustrating a macro that finds confidence intervals for the parameter estimates with a bootstrap method.

2. We thank David Riefer for emphasizing this point.

(Manuscript received August 8, 1996;  
revision accepted for publication March 5, 1997.)