Analyses of Multinomial Mixture Distributions: New Tests for Stochastic Models of Cognition and Action

Steven Yantis
Johns Hopkins University

David E. Meyer and J. E. Keith Smith
University of Michigan

Mixture distributions are formed from a weighted linear combination of 2 or more underlying basis distributions \( g(x) = \sum \omega_j f_j(x); \sum \omega_j = 1 \). They arise frequently in stochastic models of perception, cognition, and action in which a finite number of discrete internal states are entered probabilistically over a series of trials. This article reviews various distributional properties that have been examined to test for the presence of mixture distributions. A new multinomial maximum likelihood mixture (MMLM) analysis is discussed for estimating the mixing probabilities \( \omega_j \) and the basis distributions \( f_j(x) \) of a hypothesized mixture distribution. The analysis also generates a maximum likelihood goodness-of-fit statistic for testing various mixture hypotheses. Stochastic computer simulations characterize the statistical power of such tests under representative conditions. Two empirical studies of mental processes hypothesized to involve mixture distributions are summarized to illustrate applications of the MMLM analysis.

Throughout many areas of psychology, stochastic models with a combination of systematic and random components have been proposed to characterize important aspects of human behavior (e.g., Coombs, Dawes, & Tversky, 1970; Luce, Bush, & Galanter, 1963; Townsend & Ashby, 1983). The underlying components postulated by these models often correspond to a finite set of discrete probabilistic mental or physical states (e.g., levels of expectation, preparation, and motivation). Given this correspondence, a subject's observed performance (e.g., the identities, latencies, and magnitudes of overt responses) may fluctuate in a relatively simple fashion across successive trials of an experiment, depending on which state or states the subject occupies during the course of each trial.

Specifically, suppose the following:

(a) There is a random variable \( X \), the values of which represent some behavioral or psychophysiological measure (e.g., reaction time; RT) taken on a trial-by-trial basis.

(b) On each trial, the subject enters a particular state \( s_j \), selected probabilistically from a finite set of \( J \) possible states \( \{s_1, s_2, \ldots, s_J\} \).

(c) The probability of entering state \( s_j \) is \( \omega_j \), where \( 0 \leq \omega_j \leq 1 \), and \( \sum \omega_j = 1 \).

(d) The value of \( X \) obtained for a particular trial depends only on which state \( s_j \) is occupied.

(e) Each state \( s_j \) has associated with it a probability-density or probability-mass function, \( f_j(x) \), that characterizes the distribution of \( X \) when \( s_j \) is occupied.

Then we can express the distribution of \( X \) over all trials in terms of a mixture distribution that has a probability density function \( g(x) \) formed from a linear combination of the individual functions \( f_j(x) \), weighted by the probabilities \( \omega_j \) of the respective states \( s_j \). Formally,

\[
g(x) = \sum_{j=1}^{J} \omega_j f_j(x).
\]

Here the \( f_j(x) \) are called basis distributions, and the \( \omega_j \) are called mixing probabilities. In effect, the magnitudes of the mixing probabilities determine the relative contribution made by each basis distribution to the overall mixture distribution. Across the conditions of an experiment, these probabilities may change systematically. Thus, if an experiment includes \( K \) different conditions, it may yield a family of \( K \) distinct mixture distributions \( g(x) \) \((k = 1, 2, \ldots, K)\), where each member of the family is formed from the same basis distributions \( f_j(x) \) \((j = 1, 2, \ldots, J)\) but has its own unique set of mixing probabilities \( \{\omega_{k1}, \omega_{k2}, \ldots, \omega_{kJ}\} \).

An important special case of Equation 1 is the binary mixture distribution,

\[
g(x) = \alpha f_1(x) + (1 - \alpha) f_2(x).
\]

Here, \( J = 2 \) and \( \alpha = 1 - \omega_2. \) This case arises under conditions in which there are only two alternative states, \( s_1 \) and \( s_2, \)
each having its own unique basis distribution (i.e., \( f_1(x) \) and \( f_2(x) \)). For example, Figure 1 illustrates the probability-density function \( g(x) \) of a binary mixture distribution (dashed curve) formed from the density functions of two normal basis distributions (solid curves) whose means are separated by more than two standard deviation (\( \sigma \)) units.

### Stochastic Models and Mixture Distributions

Stochastic models that lead to mixture distributions have played an especially prominent role in characterizing the dynamics of human information processing (e.g., Luce, 1986; Meyer, Irwin, Osman, & Kounios, 1988; Meyer, Osman, Irwin, & Yantis, 1988; Meyer, Yantis, Osman, & Smith, 1984, 1985; Sternberg, 1973; Townsend & Ashby, 1983; Vorberg, 1981). For example, Falmagne (1965) proposed a two-state discrete stochastic model to characterize distributions of RTs from speeded choice-reaction tasks. According to this model, the observed RT on each trial comes at random from one of two alternative distributions, corresponding to two different states of subjective preparation for a presented stimulus. The subject is assumed to be prepared for some stimuli but not others. As a result, RT should be short if the subject guesses correctly and is prepared for the stimulus that actually occurs, whereas RT should be long if the subject guesses incorrectly and is not prepared. Because an all-or-none preparation process is assumed by the model, one would expect to observe a binary mixture distribution of RTs as expressed in Equation 2 (cf. Falmagne, 1965, Equation 11).

Similarly, Ollman (1966) and Yellott (1967, 1971) proposed a stochastic fast-guess model of speed-accuracy tradeoffs in choice-reaction tasks that vary the relative importance of response accuracy versus speed. This model assumes that on each trial, subjects generate either a stimulus-controlled response or a fast guess involving no stimulus information. Each of these response types is treated as having a distinct RT distribution and a distinct probability of yielding a correct response. The relative frequencies of the two response types, which reflect complementary mixing probabilities, presumably depend on incentives for speed versus accuracy. As a result, the fast-guess model yields a family of binary mixture distributions whose individual members correspond to different levels of speed stress. Yellott (1971) has evaluated the model by deriving predictions concerning the shapes of speed-accuracy tradeoff functions and comparing them with actual data.

Another interesting case comes from work by Sternberg, Knoll, and Turock (1985, 1986). Their subjects had to name visually presented target letters that appeared in linear arrays of three to six letters flashed briefly on successive trials. At a selected moment before or after the onset of each array, a visual bar marker indicated the location of the target letter to be named. When the bar marker occurred within about 150 ms after array onset, the size of the array had little effect on mean RT. However, as the delay until the bar marker increased, the array-size effect increased monotonically.

On the basis of these data, Sternberg et al. (1985, 1986) considered a two-state discrete stochastic model. According to this model, subjects have "direct access" to the visual representation of a display for a short period of time following its onset but lose such access as the representation ages, which then forces them to perform a serial search through another (nonvisual) representation. The model predicts binary mixture distributions of reaction times, resulting from fluctuations in the alternative forms of representation that subjects may use. To test the model, Sternberg et al. (1985, 1986) assessed the variances of the predicted mixtures versus the variances of the basis distributions assumed to underlie them.

Mixture distributions also arise in critical-path and order-of-processing (OP) network models (Fisher & Goldstein, 1983; Schweikert, 1982, 1983). In critical-path networks, the RT on a given trial is the sum of process durations along a critical path of the network. Because different paths are critical on different trials, the overall distribution of RTs is a mixture of the distributions for each possible critical path. Many other important cases of mixture distributions stemming from stochastic models of cognition and action could be cited as well (e.g., see Atkinson & Juola, 1974; Medin, Altom, & Murphy, 1984). Given their widespread prevalence, it is impossible for us to review them all in detail here. Clearly, however, mixture distributions have played a key role as part of psychological theory.¹

![Figure 1](image)

**Figure 1.** A mixture-distribution family involving two normal basis distributions (solid curves) and a binary mixture distribution (dashed curve) formed from them. (The mixing probability \( \alpha = 0.5 \). The larger variance, overlapping tails, multiple modes, and shared intersection point \( x_0 \) of the dashed curve relative to the solid curves illustrate several important features of binary mixtures.)

¹ Of course, there are other ways in which component process times can be combined theoretically to yield observed reaction times. One of these entails summing the individual durations of two or more processes that occur successively in time. The resultant distribution of the sum would represent a convolution of the component distributions, not a simple mixture (e.g., see Holle, 1965; Luce, 1986; Polson, Restle, & Polson, 1965; Ratcliff, 1978; Ratcliff & Murdock, 1976; Snodgrass, Luce, & Galanter, 1967). Alternatively, observed reaction times could stem from a race between underlying parallel processes, which would yield yet another form of distribution (e.g., see Meyer, Irwin, Osman, & Kounios, 1986; Townsend & Ashby, 1983). Each such possibility merits consideration in dealing with reaction-time data and other complementatory measures. For present purposes, however, we concentrate only on the analysis of mixture distributions.
Overview

The preceding examples illustrate the ubiquity of mixture distributions in models of human information processing. In this article, we introduce a new multinomial maximum likelihood mixture analysis that extends earlier work concerning mixture distributions. Our analysis is designed to provide estimates of both the basis distributions and the mixing probabilities associated with binary and higher order mixtures (i.e., those having three or more bases), given an observed set of putative mixture and basis distributions. It also yields a goodness-of-fit statistic for testing how well empirical data are approximated by mixture distributions derived from alternative finite-state stochastic models. The approach taken here is distinguished by the fact that it focuses on entire distributions, not just on a few of their parameters (e.g., mean and variance), while avoiding strong assumptions about the functional forms of the underlying bases.

The remainder of the article is organized as follows. We begin by reviewing some previous approaches to the analysis of mixture distributions, the estimation of their parameters, and the evaluation of stochastic models based on them. The advantages and disadvantages of these approaches are highlighted. Next, we briefly describe the multinomial maximum likelihood mixture analysis, including the algorithm used to implement it and a set of stochastic simulations for assessing its statistical power. Then two initial applications of this analysis to representative empirical data are summarized.

Previous Analyses of Mixture Distributions

Substantial work has been done already to explore the mathematical and statistical properties of mixture distributions (e.g., Blischke, 1964; Cox, 1966; Day, 1969; Everitt & Hand, 1981; Falmagne, 1965, 1968; Hasselblad, 1966; McLachlan & Basford, 1988; Thomas, 1969; Titterington, Smith, & Makov, 1985; Townsend & Ashby, 1983; for an elementary introduction with several examples, see Everitt, 1985). Much of this work has dealt with cases in which a mixture distribution is already known or assumed to exist, and the number and functional form of the basis distributions are also known, but the basis distributions cannot be sampled directly (Everitt & Hand, 1981). Under such circumstances, theorists have devised techniques for jointly estimating the mixing probabilities and the parameters of the basis distributions through observations sampled from the mixture distribution itself. These techniques would apply, for example, when there is a binary mixture whose underlying bases are known to be Gaussian distributions and estimates of their means and variances together with the associated mixing probabilities are desired.

In many interesting psychological applications, however, the situation is more uncertain. The number and functional form of the basis distributions may not be known. Indeed, there may be serious doubt about whether the sampled data even come from a family of mixture distributions. Under such circumstances, one would first like to test for the presence of a mixture distribution. If this test yields a positive result, then one would also like to discover how many basis distributions are involved, determine what functional form or forms they have, and estimate the values of the mixing probabilities. These objectives could, for example, play a prominent role in assessing discrete finite-state and continuous models of information-processing dynamics (Meyer, Osman, Irwin, & Yantis, 1988; Meyer et al., 1984, 1985; Yantis & Meyer, 1988).

Some progress has been made toward achieving the objectives just outlined. Mathematical and experimental psychologists have identified several quantitative properties of mixture distributions that provide useful tests without strong assumptions about the form of the underlying basis distributions (e.g., Eriksen & Yeh, 1985; Falmagne, 1968; Link, Ascher, & Heath, 1972 [cited in Lupker & Theios, 1977]; Lupker & Theios, 1975, 1977; Sternberg, 1973; Sternberg et al., 1985, 1986; Theios & Smith, 1972; Thomas, 1969; Yellott, 1971). The available tests fall roughly into two classes: ones involving the moments of mixture distributions and ones involving other special distributional features. Applications of each class have yielded informative results regarding mental processes that mediate subjects' RTs (Luce, 1986; Townsend & Ashby, 1983).

Yet many of the currently available tests have major limitations. Most notably, they often deal only with cases in which mixture distributions are binary, they do not yield estimates of the entire shapes of arbitrary basis distributions, and they do not provide statistical procedures for determining whether a mixture distribution is actually present. Such limitations are difficult to overcome for several reasons. As mentioned before, prior information about the number of possible basis distributions is often unavailable. Also, the basis distributions may overlap extensively, making it impossible to specify unambiguously which basis distribution contributed a given observation.

To illustrate the preceding points, the next two sections survey in more detail some previous tests for mixture distributions. We first consider tests based on the moments of these distributions, and then we turn to tests based on other salient distributional features.

Tests Based on Moments

Some previous tests for the presence of binary mixture distributions are based on estimating two of their moments: the mean (i.e., first moment) and the variance (i.e., second moment about the mean). In surveying these tests, we let X be a random variable that has a binary mixture distribution with mean \( \mu_x \) and variance \( \sigma_x^2 \). The means of the \( J \) underlying basis distributions are denoted \( \mu_j ( j = 1, 2) \), and the corresponding basis-distribution variances are denoted \( \sigma_j^2 ( j = 1, 2) \).

**Means of binary mixtures.** Given Equation 2, the mean of a binary mixture distribution must satisfy Equation 3:

\[
\mu_x = \alpha \mu_1 + (1 - \alpha) \mu_2,
\]

where \( \alpha \) and \( 1 - \alpha \) are the mixing probabilities associated with the first and second basis distributions, respectively (e.g., Townend & Ashby, 1983, Equation 9.6). Equation 3 therefore provides a simple and direct way to test whether a binary mixture is present. If an experimenter has an estimate of \( \mu_x \) from a sample of the hypothesized mixture distribution, and if the values of \( \alpha \), \( \mu_1 \), and \( \mu_2 \) have also been estimated through other independent procedures, then the estimate of \( \mu_x \) may be compared against the value predicted by the right side of Equation 3. The compari-
son ought to yield close agreement when a binary mixture is actually present. This might hold, for example, under conditions in which the experimenter first estimates the values of $\mu_1$ and $\mu_2$ from separate pure samples of each basis distribution and then attempts to control the value of $\alpha$ precisely while sampling the putative mixture distribution.\footnote{Equation 3 may be generalized to situations involving higher order mixture distributions. For a mixture distribution with $J$ basis distributions ($J \geq 2$), the mean $\mu_j$ is given by the following expression (Meyer, Yantis, Osman, & Smith, 1985; Townsend & Ashby, 1983): $\mu_j = \sum \sigma_j \mu_j$, where $\mu_j$ is the mean of the $j$th basis distribution ($j = 1, 2, \ldots, J$), and $\sigma_j$ is the associated mixing probability. To our knowledge, this generalization has not been used much, if at all, by psychologists in testing for the presence of higher order mixture distributions.}

An experiment by Eriksen and Yeh (1985) illustrates how Equation 3 may be used to test for the presence of a binary mixture distribution. Eriksen and Yeh studied the allocation of attention in the visual field. Subjects had to detect a designated target letter in an array of four letters. There were three different attentional pre-cueing conditions. In each condition, a pre-cue was presented at the start of a trial, indicating which of the four array locations was most likely to contain the target. The validity of the cue varied systematically across conditions, correctly specifying the target location on either 100%, 70%, or 40% of the trials. There was also a fourth “neutral” condition in which no pre-cue occurred, leaving subjects completely uncertain about the location of the impending target letter. RTs for detecting the target letters were measured as a function of the pre-cue and neutral conditions.

Eriksen and Yeh (1985) hypothesized that, on each trial, subjects would enter one or the other of two discrete attentional states: a focused-attention state in which attention was concentrated specifically on the cued location of the array or an unfocused-attention state in which attention was spread evenly over the entire array. This hypothesis implies that if a target letter actually occurs in a cued location, then the target should be detected faster when subjects are in the focused-attention state than when they are in the unfocused-attention state. Also, across trials, the RTs for the correctly cued targets should come from a binary mixture distribution, as subjects fluctuate between the focused and unfocused attention states.

To test these implications, Eriksen and Yeh (1985) used Equation 3, checking for the presence of binary mixture distributions in the RTs for correctly cued targets from the 40%-valid and the 70%-valid pre-cue conditions. The value of the mixing probability $\alpha$ under each of these conditions was assumed to equal 0.4 and 0.7, respectively, corresponding to the probability that the pre-cue correctly indicated the target’s location. In other words, subjects were assumed to “probability match,” focusing their attention on the cued location during a proportion of the trials that matched the cue validity and distributing their attention evenly on the remaining trials. The means of the hypothesized basis distributions, $\mu_1$ and $\mu_2$, which presumably reflected performance levels achieved in the focused-attention and unfocused-attention states, were estimated from the 100%-valid pre-cue and neutral conditions, respectively. This yielded reasonably accurate predictions of the mean $\mu_j$ for the 40%-valid and the 70%-valid pre-cue conditions, suggesting the presence of binary mixtures and supporting the two-state attentional hypothesis.

Other investigators have also applied Equation 3 in interesting ways. For example, Theo and Smith (1972) used it to evaluate predictions derived from Falmagne’s (1965) discrete two-state preparation model. Consistent with the model, they found that binary mixture distributions of RTs appeared to be present across a variety of conditions in a sequential choice-reaction task.

As the applications discussed thus far illustrate, sample means can provide a test for binary mixture distributions without requiring assumptions about the functional forms of the underlying basis distributions. It should be noted, however, that this sort of test is quite weak. Equation 3 has a high likelihood of yielding Type II errors (i.e., incorrectly accepting the null hypothesis that a binary mixture distribution is present). Under experimental conditions in which a binary mixture distribution is not present, the sample mean may still easily fall between the means from other conditions thought to manifest the basis distributions individually (Meyer et al., 1985; Townsend & Ashby, 1983). As we show later, one can obtain more powerful tests by looking not only at means but also at other aspects of data where mixtures might be present.

Proportional speed–accuracy tradeoff. Yellott (1971), for instance, derived a useful extension of Equation 3. In experiments that systematically varied the speed and accuracy of subjects’ performance, he showed how the presence of binary mixture distributions can be tested by examining a combination of mean RTs and error rates. His test stems from the fast-guess model outlined previously (Ollman, 1966; Yellott, 1967, 1971), which assumes that performance involves a discrete two-state response process. On each trial, the subject is assumed to make either a slow and typically correct stimulus-controlled response, or a fast and frequently incorrect guessing response. According to this model, the resultant RTs should have a binary mixture distribution composed of contributions from the two different response states (i.e., stimulus controlled and fast guess), and the following proportional speed-accuracy tradeoff relation should hold:

$$P_x \mu_e - P_x \mu_f = \mu_x (P_e - P_x). \tag{4}$$

Here $P_x$ is the overall probability of a correct response, $\mu_x$ is the mean of the RT distribution for correct responses, $P_e$ is the overall probability of an incorrect (error) response (i.e., $P_e = 1 - P_x$), $\mu_e$ is the mean of the RT distribution for incorrect responses, and $\mu_f$ is the mean of the basis RT distribution for stimulus-controlled responses (Yellott, 1971).

Equation 4 has an interesting form. It implies that the difference between the mean RTs of the correct and incorrect responses weighted by their respective probabilities of occurrence (i.e., $P_x \mu_e - P_x \mu_f$) will be directly proportional to the difference between these probabilities themselves (i.e., $P_e - P_x$), where the multiplicative proportionality constant is $\mu_x$. One may test whether a set of data satisfies this implication without observing the basis (i.e., stimulus controlled and fast guess) RT distributions directly, because the test merely requires values of $P_x$, $\mu_x$, $P_e$, and $\mu_f$, which are all directly estimable from each of several different speed-accuracy tradeoff conditions. If the manipulation of subjects’ speed and accuracy yields a proportional
tradeoff relation (i.e., Equation 4), then one would have evidence for a family of binary mixture distributions, corresponding to various mixtures of the two putative response states. In particular, Yellott (1971) found good agreement between Equation 4 and data from studies of two-choice RTs.

The approach taken by Yellott (1971) has some notable virtues. It seems less likely that Equation 4 would falsely suggest the presence of binary mixture distributions than Equation 3 would. This is because Equation 4 requires the data to have a rather simple but unique pattern: direct proportionality between RT differences and accuracy differences across conditions. Presumably, such a pattern could not arise easily from sources other than binary mixtures. Yet the circumstances under which one may apply Equation 4 are limited. They require inducing subjects to vary their response speed and accuracy systematically. They also require assuming that the fast-guess model (Yellott, 1971) is correct. In the absence of these particular manipulations and assumptions, additional tests for the presence of mixture distributions are needed.

**Variances of binary mixtures.** Another useful test for the presence of binary mixture distributions involves their variances. The variance of a binary mixture distribution, \(\sigma^2\), should satisfy Equation 5 (e.g., Falmagne, Cohen, & Dwivedi, 1975, p. 371; Townsend & Ashby, 1983, Equation 9.7):

\[
\sigma^2 = \alpha \sigma_1^2 + (1 - \alpha) \sigma_2^2 + \alpha(1 - \alpha)(\mu_1 - \mu_2)^2,
\]

where \(\sigma_1^2\) and \(\sigma_2^2\) are the variances of the two basis distributions, \(\mu_1\) and \(\mu_2\) are the corresponding means, and \(\alpha\) is the mixing probability. Consequently, if one has separate estimates of \(\sigma_1^2\), \(\sigma_2^2\), \(\mu_1\), and \(\mu_2\) obtained from pure samples of each basis distribution, and if one knows the value of the mixing probability \(\alpha\), then one can predict the expected value of the variance for a hypothesized binary mixture when the bases are combined. Comparing this prediction with observed data provides a test of whether the mixture is actually present.3

One might anticipate, in particular, that the variance of a binary mixture distribution would be larger than the variances of the contributing basis distributions. Equation 5 contains a linear combination of the basis distributions' variances weighted proportionally by the mixing probabilities (i.e., \(\alpha \sigma_1^2 + (1 - \alpha) \sigma_2^2\)), plus an inflation factor consisting of the squared difference between the basis distributions' means weighted by the product of the mixing probabilities (i.e., \(\alpha(1 - \alpha)(\mu_1 - \mu_2)^2\)). When the inflation factor is large relative to the variances of the basis distributions, the variance of the binary mixture distribution will exceed them. This can happen most readily if the mixing probabilities have intermediate values (i.e., \(\alpha = 1 - \alpha = 0.5\)) and the squared difference between the means of the basis distributions (i.e., \((\mu_1 - \mu_2)^2\)) dominates the variance of each basis distribution. For example, these circumstances are illustrated in Figure 1, where the mixture distribution (dashed curve) has a much greater variance than do the associated basis distributions (solid curves).

Given the potential utility of Equation 5, several experimenters have applied it in testing for the presence of binary mixture distributions (e.g., Eriksen & Eriksen, 1972; Meyer et al., 1983; Sternberg et al., 1986). For example, Eriksen and Eriksen (1972) studied the effect of backward masking on visual stimulus identification. During each trial of their experiment, a printed letter was presented briefly. Following a set interstimulus interval (ISI) after the onset of the test stimulus, a subsequent masking stimulus was presented. Subjects' RTs and response accuracy in identifying the test stimuli were measured as a function of the ISI. The variances of the RT data provided evidence to evaluate a discrete two-state model of stimulus identification that implies the presence of binary mixture distributions.

Eriksen and Eriksen (1972) hypothesized specifically that during each trial of their experiment, subjects' responses were based on one or the other of two possible perceptual states: a *degraded-stimulus state* in which the test stimulus was not perceived clearly before the subsequent masking stimulus obliterated it and an *intact-stimulus state* in which the test stimulus was perceived clearly. According to this hypothesis, the probability of the intact-stimulus state should increase monotonically with the length of the ISI, yielding a monotonic increase in the probability of correct responses. Also, for intermediate ISIs, the observed RTs should come from binary mixture distributions, corresponding to combinations of samples generated stochastically from the two alternative perceptual states.

To test this latter prediction, the variances of the RTs were examined as a function of the ISI duration. Eriksen and Eriksen (1972) expected that if the RTs really came from a family of binary mixture distributions, their variances should first increase and then decrease as the duration of the ISI increased, yielding an inverted-U function. This expectation follows directly from Equation 5, whose inflation factor \(\alpha(1 - \alpha)(\mu_1 - \mu_2)^2\) is largest for intermediate values of the mixing probability \(\alpha\), corresponding to intermediate ISI durations. However, the observed data revealed a monotonically decreasing function of RT variance versus ISI. Eriksen and Eriksen (1972) therefore rejected a discrete two-state (all-or-none) model of stimulus identification in favor of one that assumes the perceptual clarity of a stimulus increases gradually throughout the ISI.

A related test for the presence of binary mixture distributions was proposed by Falmagne et al. (1975; also see Townsend & Ashby, 1983, pp. 264–265). They noted that the variances of such distributions should satisfy Equation 6:

\[
\sigma^2 = -\mu_x^2 + A \mu_x + B,
\]

where \(\sigma^2\) and \(\mu_x\), respectively, are the variance and mean of the mixture distribution, and \(A\) and \(B\) are constants whose values depend only on the means and variances of the underlying basis distributions (i.e., \(\mu_1, \mu_2, \sigma_1^2\), and \(\sigma_2^2\)). This result follows directly from combining Equations 3 and 5. Equation 6 has the advant-

3 Like Equation 3, Equation 5 may be generalized to situations involving higher-order mixture distributions. For a mixture of \(J\) basis distributions \((J \geq 2)\), the variance \(\sigma^2\) is given by the following expression (Meyer, Yantis, Osman, & Smith, 1985):

\[
\sigma^2 = \sum_{j=1}^{J} \alpha_j \sigma_j^2 + \sum_{j=1}^{J} \alpha_j (\mu_j - \mu_x)^2,
\]

where \(\mu_j\) is the mean of the \(j\)th basis distribution \((j = 1, 2, \ldots, J)\), \(\sigma_j^2\) is the variance of the \(j\)th basis distribution, \(\alpha_j\) is the \(j\)th mixing probability, and \(\mu_x\) is the mean of the binary mixture distribution (cf. Footnote 2).
tage that it may be applied without estimating the mixing probability \( \alpha \). Across a set of experimental conditions that involve the same basis distributions mixed in various proportions, Equation 6 must hold under each condition, regardless of the value of the mixing probability.

Variance tests for the presence of binary mixture distributions therefore significantly augment those based solely on means. By applying Equations 5 and 6 to compare predicted variances with observed variances, one can at least partly take into account some major features of the shapes of these distributions. This account can be developed without knowing the exact functional forms of the underlying basis distributions or the values of the mixing probabilities associated with them.

Nevertheless, the variance tests still leave much to be desired. They will not work under all conditions. Some conceivable basis distributions (e.g., Cauchy) have shapes that are difficult to distinguish from normal (Gaussian) distributions, but their means do not satisfy the central limit theorem, and their variances do not have finite values. For mixture distributions formed from such bases, attempts to apply Equations 5 and 6 could yield misleading results.

A similar problem can arise even when the basis distributions are not as pathological as this. If they have heavily skewed tails, or if the data sampled from them contain artificial outliers, then an experimenter would encounter difficulty in estimating the required variances accurately. Very large sample sizes may be necessary to overcome such difficulties, and this in turn may compromise the stationarity of the sampling process. Consequently, additional tests beyond those involving the moments of mixture distributions and their bases are needed.

**Moment-Free Tests**

Fortunately, mathematical psychologists have developed some tests that do not require estimates of distributional moments (i.e., means, variances, etc.) to detect the presence of mixture distributions. We call these moment-free tests. They concern special features of binary mixtures revealed by the shapes of empirical relative frequency and cumulative distribution functions (e.g., Falmagne, 1968; Link et al., 1972 [cited in Lupker & Theios, 1977]; Lupker & Theios, 1975, 1977; Sternberg, 1973; Thomas, 1969; Townsend & Ashby, 1983). In many cases, tests involving such features may be applied without having information about the extreme tails of the underlying basis distributions.

**Bimodality.** For example, one simple and popular moment-free test involves bimodality (Townsend & Ashby, 1983). If a putative mixture distribution has two underlying unimodal basis distributions, and if the basis distributions are far enough apart, then the probability-density function of the mixture (Equation 2) should have two distinct modes, as illustrated in Figure 1 (dashed curve). An observation of bimodality in an empirical frequency distribution may therefore be taken as tentative evidence that a binary mixture distribution is present.

However, binary mixture distributions need not be bimodal, even if their basis distributions are unimodal (e.g., see Figure 3). In particular, suppose that the basis distributions \( f_i(x) \) and \( f_2(x) \) have normal (Gaussian) density functions with the same variance \( N(\mu_i, \sigma^2) \), \( i = 1, 2 \) and \( \alpha = 0.5 \). Then the mixture density \( g(x) = 0.5 f_1(x) + 0.5 f_2(x) \) will have two modes if and only if \( |\mu_1 - \mu_2| > 2\sigma \) (Everitt & Hand, 1981, Equation 2.4). Furthermore, sampled distributions that are not binary mixtures (particularly one whose measurement is subject to experimental error) can still be bimodal. Thus, strictly speaking, bimodality is neither necessary nor sufficient evidence for a binary mixture.

**Fixed-point property.** A potentially more informative moment-free test has been developed by Falmagne (1968). He noted that for binary mixtures (Equation 2), if there is a point \( x_0 \) where the two basis density functions cross (i.e., \( f_1(x_0) = f_2(x_0) = \phi \)), then all mixture density functions \( g_\alpha(x) \) formed exclusively from them (i.e., \( g_\alpha(x) = \alpha_1 f_1(x) + (1 - \alpha_1) f_2(x) \)) must also cross there (i.e., \( g_\alpha(x_0) = \phi \vee \alpha_1 \)). This feature of binary mixtures, known as the fixed-point property (Falmagne, 1968), is illustrated in Figure 1, where the mixture density (dashed curve) and basis densities (solid curves) cross at the intermediate point \( x_0 \).

The fixed-point property has been used repeatedly to test for binary mixture distributions in RT data. In analyzing RTs from a six-choice reaction task, Falmagne (1968) found that several empirical distributions sampled from putative binary mixtures that supposedly had the same two basis distributions did not exhibit a fixed crossover point. He thus concluded that either (a) a mixture with at least three bases was involved or (b) there were no mixture distributions at all in his experimental situation.

Subsequently, Lupker and Theios (1975, 1977) suggested that the RTs from Falmagne's six-choice reaction task may indeed have embodied the contributions of three or more basis distributions, corresponding to three distinct subjective preparatory states (i.e., extremely unprepared, neutral, and prepared). This suggestion differed from Falmagne's (1965) original hypothesis, which was based on a model having only two preparatory states (i.e., unprepared and prepared). The existence of a third, extremely unprepared, state seemed plausible to Lupker and Theios (1975, 1977) because some of Falmagne's (1965) stimulus probabilities were exceptionally low (i.e., \( \leq 0.06 \)). In their own experiments, they always had larger stimulus probabilities to eliminate the extremely unprepared state and to encourage performance based on only two preparatory states. Under these new conditions, results more consistent with the fixed-point property of binary mixture distributions were obtained.

As the preceding examples illustrate, the fixed-point property is useful when the exact forms of the two basis distributions are not known and cannot be observed directly, but samples from two or more hypothesized binary mixture distributions are available. However, if the forms of the bases are known or can be observed directly, then tests that incorporate the full power of Equation 1 are preferable. This is because the fixed-point property involves just a single point from any one distribution, rather than taking entire distribution shapes into account. Also, the fixed-point property only applies to cases in which all the putative basis density functions cross at a single point. Although this condition is certain to be met by all binary mixtures whose basis distributions overlap at all, it is very unlikely to be met by mixtures involving more than two basis distributions.

**Compound-mixture property.** One moment-free test that
does deal with entire distribution shapes in testing for a binary mixture has been proposed by Thomas (1969). He noted that three different mixtures \( g_1(x), g_2(x), \) and \( g_3(x) \) of two underlying basis distributions must themselves be mixtures of one another, as expressed in Equation 7:

\[
g_k(x) = p g_1(x) + q g_2(x),
\]

where the \( g_k(x) (k = 1, 2, 3) \) have associated mixing probabilities \( \alpha_k \) that satisfy Equation 2 and \( p \) and \( q \) are linear combinations of the \( \alpha_k \) and their equations. Equation 7 is known as the compound-mixture property (cf. Sternberg et al., 1986, Appendix 1).

It is especially relevant when direct estimates of the underlying basis distributions \( \{f_1(x) \text{ and } f_2(x)\} \) are not available, but an investigator has estimates of three or more mixture distributions formed from these bases. By testing whether one putative mixture \( \{g_1(x) \text{ and } g_2(x)\} \) is a linear combination of the other two \( \{g_1(x) \text{ and } g_3(x)\} \), as Equation 7 specifies, one may assess the credibility of the hypothesized mixtures.

For obvious reasons, however, this test still has serious limitations. It does not take into account the shapes of the individual basis distributions per se. When they are directly estimable from existing data or known already through other means, then further tests involving the shape of each basis distribution may be preferable.

**Stochastic dominance relations.** Some further moment-free tests that do take the shapes of basis distributions more fully into account may be formulated through stochastic dominance relations. A stochastic dominance relation concerns the relative magnitudes of two different cumulative distribution functions as they range from zero to one. To be specific, suppose that \( F_1(x) \) and \( F_2(x) \) are the respective cumulative distribution functions (CDFs) of two random variables \( X_1 \) and \( X_2 \), where \( F_1(x) = P(X_1 \leq x) \) and \( F_2(x) = P(X_2 \leq x) \). Then \( F_1(x) \) stochastically dominates \( F_2(x) \) whenever

\[
F_1(x) \geq F_2(x) \quad \forall x.
\]

In this case, \( F_1(x) \) dominates \( F_2(x) \) because \( P(X_1 \leq x) \) always equals or exceeds \( P(X_2 \leq x) \).

More generally, stochastic dominance relations may hold among an ordered set of \( J \) cumulative distribution functions \( \{F_j(x) : j = 1, 2, \ldots, J\} \). When they do, the CDFs will have certain interesting features. Suppose, in particular, that for all pairs of CDFs \( F_j(x) \) and \( F_{j'}(x) \), \( F_j(x) \) dominates \( F_{j'}(x) \) if and only if \( j < j' \). Then the \( J \) distribution functions would never cross each other, and no two of their probability-density functions \( \{f_j(x), f_{j'}(x), \ldots, f_J(x)\} \) would ever cross more than once. For example, the two solid curves in Figure 2 [labeled \( F_1(x) \) and \( F_2(x) \)] illustrate a set of ordered CDFs that satisfies the stochastic dominance relation expressed by Equation 8.

Using such stochastic dominance relations, Sternberg (1973) identified two complementary properties that the shapes of mixture distributions and their basis distributions should exhibit in certain studies of reaction time. The first is called the short-RT property, which states that a mixture of \( J \) basis distributions ordered in terms of Equation 8 should yield some observations that are as small as the smallest ones obtained from the basis distribution with the most dominant CDF. More precisely, suppose that \( F_1(x) \) is the most dominant CDF and that

\[
G(x) \text{ is the CDF of a mixture distribution to which } F_1(x) \text{ contributes. Then } G(x) \geq \alpha F_1(x) \quad \forall x, \text{ where } \alpha \text{ is the mixing probability} \ (0 < \alpha < 1) \text{ associated with } F_1(x). \text{ This is illustrated by the lower left end of the dashed curve in Figure 2.}
\]

Sternberg's (1973) short-RT property stems directly from the definition of a mixture distribution (Equation 1) and the definition of a stochastic dominance relation (Equation 8). Based on these equations,

\[
G(x) = \sum_{j=1}^{J} \alpha_j F_j(x).
\]

In turn, Equation 9 implies that \( G(x) \geq \alpha F_1(x) \quad \forall x, \) because all the remaining terms in the sum \( (j > 1) \) are nonnegative. Also, because \( \alpha > 0 \) and \( F_1(x) \geq F_{j'}(x) \quad \forall j' \geq 1 \) (Equation 8), the point at which \( G(x) \) starts to exceed zero must be the same as the point at which \( F_1(x) \) starts to exceed zero. More precisely, \( \text{glb}[xG(x) > 0] = \text{glb}[xF_1(x) > 0] \), where \( \text{glb} \) denotes greatest lower bound. Thus, the mixture distribution should yield some observed values of \( x \) that are as small as the smallest ones obtained from the most dominant basis distribution.

A second feature identified by Sternberg (1973) to characterize mixture distributions is the long-RT property. It states that a mixture of \( J \) basis distributions ordered in terms of Equation 8 should yield some observations as large as the largest ones obtained from the basis distribution with the least dominant CDF. More precisely, suppose that \( F_L(x) \) is the least dominant CDF and that \( G(x) \) is the CDF of a mixture distribution to which \( F_L(x) \) contributes. Then \( 1 - G(x) \geq \alpha_L[1 - F_L(x)] \), where

\footnote{Parallel to the work of Thomas (1969), a similar test has been developed by Sternberg, Knoll, and Turok (1986, Appendix 1).}
\[ a_j \] is the mixing probability \((0 < a_j < 1)\) associated with \(F_j(x)\). This is illustrated by the upper right end of the dashed curve [i.e., \(G(x)\)] in Figure 2.

Like the short-RT property, Sternberg's (1973) long-RT property stems directly from the definition of a mixture distribution (Equation 1) and the definition of a stochastic dominance relation (Equation 8). Again, based on these equations,

\[ 1 - G(x) = 1 - \sum_{j=1}^{J} a_j F_j(x) = \sum_{j=1}^{J} a_j [1 - F_j(x)]. \quad (10) \]

In turn, Equation 10 implies that \(1 - G(x) \geq a_j[1 - F_j(x)]\), because all of the remaining terms in the sum \((j < J)\) are nonnegative. Also, because \(a_j > 0\) and \(1 - F_j(x) \geq 1 - F_j(x) \forall j < J\) (Equation 8), the point at which \(G(x)\) first reaches \(1\) [and \(1 - G(x)\) first reaches \(0\)] must be the same as the point at which \(F_j(x)\) first reaches \(1\) [and \(1 - F_j(x)\) first reaches \(0\)]. In other words, \(\text{lub}[x|G(x) < 1] = \text{lub}[x|F_j(x) < 1]\), where \(\text{lub}\) denotes least upper bound. Thus, the mixture distribution should yield some observed values of \(x\) that are as large as the largest ones obtained from the least dominant basis distribution.

An appealing feature of the short-RT and long-RT properties is that they may be used to test for the presence of both binary and higher order mixture distributions. The test requires a specification of the assumed dominance relations among the putative basis distributions, but it does not require knowing the exact forms of those distributions. Whenever a putative mixture comes from sampling a set of ordered basis distributions in which the stochastic dominance relation (Equation 8) holds, results consistent with the two properties should emerge. Failure to obtain such results would constitute evidence against the hypothesis that a mixture distribution is present.

Sternberg (1973) noted, for example, that self-terminating search processes imply higher order mixture distributions of RTs having the short-RT and long-RT properties. However, his own data concerning these processes did not exhibit the predicted properties, so he ultimately questioned the notion of self-terminating search. Additional applications of the short-RT property may be found in Sternberg and Knoll (1986) and Sternberg, Knoll, and Turock (1990). Vorber (1981) further generalized Sternberg's (1973) analysis of the short-RT and long-RT properties.

There are also other useful moment-free tests for binary mixture distributions based on stochastic dominance relations. One of these involves what we call the homogeneous-difference property of mixture CDFs. As Lupker and Theios (1977) have noted, the CDFs of different binary mixtures [i.e., \(G_a(x)\) and \(G_b(x)\)] formed from the same two basis distributions may be written as

\[ G_a(x) = a_F(x) + (1 - a_b) F_2(x) \quad (11) \]

and

\[ G_b(x) = a_b F_1(x) + (1 - a_b) F_2(x). \quad (12) \]

From these equations, we obtain

\[ G_a(x) - G_b(x) = [a_b - a_F] [F_1(x) - F_2(x)]. \quad (13) \]

Thus, if \(F_1(x)\) stochastically dominates \(F_2(x)\), then \(G_a(x) - G_b(x)\) must be uniformly nonnegative whenever \(a_F \geq a_b\), and uniformly negative otherwise. For fixed mixing probabilities \(a_F\) and \(a_b\), a stochastic dominance relation between two basis distributions therefore implies that the CDFs of any two binary mixture distributions formed from them will never cross. So like Sternberg's (1973) short-RT and long-RT properties, the homogeneous-difference property (Equation 13) and its consequences may be checked to test for the presence of such mixtures.

In addition, binary mixture distributions should exhibit a simple characteristic that we call the constant-ratio property (cf. Link et al., 1972 [cited in Lupker & Theios, 1977]; Lupker & Theios, 1977). This property concerns the ratio of differences between pairs of CDFs. Suppose that \(F_1(x)\) and \(F_2(x)\) are the respective CDFs of two basis distributions, as in our previous discussion (Equations 11 to 13). Suppose further that \(G_a(x)\) and \(G_b(x)\) are the CDFs of mixture distributions formed from these basis distributions according to Equations 11 and 12, respectively, and \(G_a(x)\) is the CDF of a third binary mixture such that \(G_a(x) = a_F F_1(x) + (1 - a_b) F_2(x)\). Then the difference between \(G_a(x)\) and \(G_b(x)\) may be expressed as

\[ G_a(x) - G_b(x) = [a_b - a_F] [F_1(x) - F_2(x)]. \quad (14) \]

Dividing each side of Equation 14 into the corresponding side of Equation 13 yields

\[ \frac{G_a(x) - G_b(x)}{[G_a(x) - G_b(x)]} = \frac{[a_b - a_F] [F_1(x) - F_2(x)]}{[a_b - a_F] [F_1(x) - F_2(x)]} = \frac{a_b - a_F}{a_b - a_F}. \quad (15) \]

Thus, the ratio of the differences between pairs of binary mixture CDFs should remain constant as \(x\) increases, corresponding to the ratio of the differences between their mixing probabilities.

In an analysis of RT data from four-choice and six-choice reaction tasks, Lupker and Theios (1977) checked for the presence of binary mixture distributions by applying a battery of moment-free tests such as those outlined here. The data satisfied the constant-ratio property (Equation 15) of Link et al. (1972; cited in Lupker & Theios, 1977) reasonably well. Sternberg's (1973) short-RT and long-RT properties, along with the homogeneous-difference property of mixture CDFs (Equation 13), were not violated either. Similarly, a reasonable fit was obtained for Falmagne's (1968) fixed-point property. Together, these results illustrate how one moment-free test may reinforce conclusions drawn from others, thereby helping overcome the individual weaknesses of each test.

Summary

The moment-free tests for mixture distributions reviewed here (e.g., fixed-point property, homogeneous-difference property, and constant-ratio property) provide a useful complement to those based on distributional moments (i.e., mixture means and variances). It should be emphasized, however, that these tests still leave much to be desired. Their limitations include the following:

1. Many moment-free tests apply mainly to situations involv-
ing binary mixture distributions, and not to higher order mix-
tures involving more than two basis distributions.

2. Although these tests focus on some detailed features of
the probability-density and cumulative distribution functions
associated with mixture distributions and their bases, they do
not typically yield precise estimates of each individual function
per se.

3. Goodness-of-fit statistics that quantify how closely a set of
data corresponds to a specified mixture distribution are often
lacking. Investigators may instead have to rely on subjective
impressions to determine whether a hypothesized mixture dis-
tribution is really present.

Multinomial Maximum Likelihood Mixture Analysis

Given the limitations of moment-free tests and the problems
with tests that rely on distributional moments, we have de-
veloped another approach toward checking for the presence
of mixture distributions (Smith, Meyer, Yantis, & Osman, 1982).
Our approach involves a new multinomial maximum likelihood
mixture analysis (MMLM analysis). In applications of this an-
alysis, separate samples of observations must be drawn from each
of two or more putative basis distributions and from the mix-
ture distributions to which they are thought to contribute. After
these samples are obtained, the probability-density functions of
both the mixture distributions and their underlying basis distri-
butions may be estimated. Furthermore, the MMLM analysis
yields estimates of the mixing probabilities associated with the
mixture and a goodness-of-fit statistic that quantifies how well
a hypothesized mixture actually accounts for the data.

The MMLM analysis has several advantages. It takes the en-
tire shapes of mixture distributions and their basis distribu-
tions into account, but it does not require any knowledge or
assumptions about the exact functional forms of these distribu-
tions. For example, the basis distributions might come from a
Gaussian, gamma, or exponential distribution family, but no
such family has to be specified as part of the MMLM analysis.
Also, the analysis is not limited to situations involving binary
mixtures; it applies to higher order mixtures (i.e., those having
more than two bases) as well.

The following sections of the article outline the MMLM anal-
ysis in schematic form; a more detailed description of the anal-
ysis and its mathematical derivation may be found in Smith,
Meyer, and Yantis (1991). Our presentation begins with a geo-
metric characterization of how the analysis works. Next we de-
scribe parameter estimation and the statistical tests based on it,
and we assess the power of these tests. Finally, we review some
representative studies that have applied the MMLM analysis to
investigate the dynamics of human information processing.5

Data Structures and Theoretical Assumptions

Empirical observations. In applications of the MMLM anal-
ysis, data are first collected from a set of $K$ distinct experimen-
tal conditions $\{C_i; k = 1, 2, \ldots, K\}$. Each condition is arranged
to yield a set of $n_k$ empirical observations $\{x_{nk}; n = 1, 2, \ldots, n_k\}$
representing sampled values of some continuous random vari-
able $X$ (e.g., reaction time). The observations from the $k$ th con-
dition are assumed to be stochastically independent, identically
distributed, and representative of an underlying theoretical dis-
tribution whose probability-density function is $f_k(x)$. Also, for
present purposes, it is assumed that the observations from con-
dition $C_i$ are stochastically independent of those from con-
dition $C_j$, for all $k \neq j$.

These stochastic-independence assumptions enable us to im-
plement a relatively simple form of maximum likelihood esti-
mation for the MMLM analysis. If these assumptions are vi-
olated, the parameter estimates and goodness-of-fit statistics
provided by the analysis could be systematically biased. How-
ever, it is essentially impossible to quantify the potential mag-
nitude of such biases without a precise model for expected inter-
observation dependencies, which we do not have at the present
time. Furthermore, we cannot say for sure whether the depen-
dencies would make the analysis more or less likely to reject a
mixture-distribution hypothesis. What can be said is that inves-
tigators who wish to use the analysis should, in their experi-
ments, minimize the degree of stochastic dependence between
observations by implementing appropriate control procedures.
These would include such standard steps as maintaining sub-
jects' level of alertness at a constant high level across trials,
providing adequate recovery time between trials, limiting the
total number of trials per session, and eliminating contingen-
cies between successive stimulus presentations as much as pos-
able.

Partitioning of conditions and distributions. A major ob-
jective of the MMLM analysis is to test the hypothesis that the
theoretical distributions associated with the $K$ experimental
conditions form a family of basis and mixture distributions.
According to this hypothesis, there are $J$ conditions in the over-
all set ($J < K$) each of which has a unique pure basis distribu-
tion associated with it, and $K - J$ conditions that are mixtures
of these bases. For present purposes, the $K$ conditions are part-
tioned into two sets: A set of $J$ conditions that are identified as
basis conditions, and a set of $K - J$ conditions that are identi-
fied as mixture conditions. We refer to the $J$ basis conditions as
$C_j$ and to the density functions that involve pure nominal basis
distributions as $f_j(x)$ ($j = 1, 2, \ldots, J$). The distributions for
the remaining $K - J$ conditions $\{C_k; k = J + 1, \ldots, K\}$ are
hypothesized to be distinct mixtures of the basis distributions
associated with the $J$ basis conditions; that is, $f_k(x) = \sum_{j=1}^{J} \rho_{jk} f_j(x)$, with each mixture distribution having its own
unique set of mixing probabilities $\{\rho_{jk}; j = 1, 2, \ldots, J\}$.

At this point, no additional formal assumptions are required
about the relations among the putative basis distributions. Ordin-
arly, one might expect them to differ significantly. It is conceiv-
able, however, that one or more of the bases could themselves
be linear combinations (e.g., mixtures) of the others. If this were
the case, then most of the mathematical development pre-
sented here would still hold. The only marked consequence of

5 A program that implements the MMLM analysis may be obtained
from the authors. Send requests for the program to Steven Yantis, De-
partment of Psychology, Johns Hopkins University, Baltimore, Mary-
land 21218. Enclose either an IBM-compatible computer diskette and
appropriate self-addressed packaging with return postage, or $10.00 to
cover same (specifying diskette size—5.25" or 3.5"). In return, you will
receive a diskette containing the commented source code in Turbo
Pascal, the executable code, and some sample input and output files.
having linear dependencies among some of the basis distributions would be an increase in the sizes of the confidence intervals surrounding the estimated mixing probabilities assigned to these distributions. 

On the other hand, the requirement that we first partition the experimental conditions into those that are identified with pure basis distributions and those that involve mixtures of the bases imposes a relatively strong constraint. Unlike the MMLM analysis, most previous treatments of mixture distributions have weaker constraints in this regard. As mentioned previously, some of the mixture-distribution tests that examine distributional moments (e.g., Equations 4 and 6) or moment-free features (e.g., the compound-mixture, homogenous-difference, and constant-ratio properties) can be applied without any pure samples of the underlying basis distributions. What they lose as a result, however, is the extra power provided by empirical evidence about the overall shapes of the individual basis distributions, which the MMLM analysis specifically takes into account.

Illustrative example. An illustration of one simple case to which the MMLM analysis would apply when we have partitioned the experimental conditions appears in Figure 3. This figure shows three hypothetical density functions (solid curves) for a random variable X. In the top and bottom panels of the figure, the density functions represent two putative basis distributions, \( f_1(x) \) and \( f_2(x) \), associated with two basis conditions. In the middle panel is the density function for a mixture distribution, \( f_2(x) \), formed from the two basis distributions by mixing them equally (i.e., \( f_2(x) = \alpha f_1(x) + (1 - \alpha) f_2(x); \alpha = 0.5 \)). Above the horizontal axis of each panel, the small dots mark representative numerical observations that one might obtain from sampling the density functions randomly. For this case, which involves a binary mixture distribution family (Equation 2), we have \( J = 2, K = 3, \) and \( n_k = 20 \) in each condition \((k = 1, 2, 3)\).

Given cases like the one in Figure 3, as well as more complex ones with additional experimental conditions \((K > 3)\) and additional basis distributions \((J > 2)\), the MMLM analysis has three straightforward objectives: (a) estimate the bin probabilities for the basis conditions, (b) estimate the mixing probabilities for the mixture conditions, and (c) assess how well the estimated distributions account for the observations obtained within and across conditions. Achieving these objectives, however, is not completely straightforward. Although the experimental conditions and basis distributions are assumed to be correctly partitioned (i.e., one knows a priori which conditions have pure basis distributions associated with them), we make no assumptions about the particular shapes of the basis distributions. This lack of constraint complicates the estimation problem. The task is also complicated because we only have a finite number of observations \((n_k)\) in each condition, which blurs our view of the distributions' individual shapes to some extent.

Grouping into bins. To deal with these complications while achieving its stated objectives, the MMLM analysis condenses quantitative information in the available observations through a categorical data analysis procedure (cf. Hall & Titterington, 1984). We begin by grouping the observations \((x_{nk}; k = 1, 2, \ldots, K; n = 1, 2, \ldots, n_k)\) obtained from the \( K \) experimental conditions into a set of \( J \) subintervals or bins, whose boundaries are the same for each sampled basis and mixture distribution, as Figure 3 illustrates (dashed vertical lines). For technical reasons (see later discussion of Equation 20), the number of bins must exceed the number of hypothesized basis distributions (i.e., \( J > K \)). The number and placement of the bin boundaries are also chosen such that, typically, each condition contributes on the order of at least five observations per bin (cf. Hays, 1981).

After the observations have been grouped into bins, the frequencies of observations within each bin are calculated as a function of the experimental condition from which they came. This yields an empirical frequency matrix \( N \) with \( J \) rows and \( K \) columns, where the rows and columns correspond respectively to the selected bins \((B; i = 1, 2, \ldots, J)\) and experimental conditions \((C_k; k = 1, 2, \ldots, K)\). Each entry \( n_{ik} \) of \( N \) represents the number of observations contributed by the \( k \)th condition and its associated distribution \( f_k(x) \) to the \( i \)th bin. Within any column of \( N, \ n_{ik} = \Sigma n_{ik}, \) the total number of observations obtained in condition \( C_k \). For example, Table 1 illustrates the frequency matrix \( N \) derived from grouping the marked observations (small dots) of Figure 3 into 6 bins; here, \( n_1 = n_2 = n_3 = 20 \).

The observation frequencies from the various experimental conditions may also be represented graphically in terms of \( K \) histograms. For each condition \( C_k \) and each column of \( N \), there is a distinct histogram whose shape more or less closely approximates that of the underlying theoretical density function \( f_k(x) \), with the degree of approximation depending on the observation sample size \( n_k \) and the width of the bins. This approximation is illustrated in Figure 4, which shows a set of histograms obtained from the cell entries \( n_{ik} \) of the frequency matrix \( N \) (Table 1; cf. Figure 3).

Alternatively, the columns of the empirical frequency matrix \( N \) can be viewed as representing a set of \( K \) vectors in an \( L \)-dimensional space. Here, for each experimental condition \( C_k \) \((k = 1, 2, \ldots, K)\), the associated frequency vector is \( \mathbf{n}_k = \{n_{1k}, n_{2k}, \ldots, n_{jk}\} \). As we show later, this vector representation may significantly augment the histogram representation (Figure 4) as part of the MMLM analysis.

Of course, it should be recognized that grouping the observations from each experimental condition into a finite set of bins has both costs and benefits. A potentially significant cost is that we lose some quantitative information about the exact shapes of the underlying basis and mixture distributions. In effect, the grouping procedure functions as a low-pass filter, suppressing high-fidelity details of the distribution shapes. The effects of this suppression can be overcome at least partially by increasing the number of bins and the sample size in each condition, while decreasing the bin widths. More important, however, observation grouping has a crucial benefit: It allows us to formulate a principled set of statistical techniques, including the foundation for a quantitative goodness-of-fit test.

Relation to theoretical parameters. Once the empirical frequency matrix \( N \) has been derived for a set of observations from the \( K \) experimental conditions, the next step of the MMLM analysis is to relate this matrix to theoretical parameters of the hypothesized basis and mixture distributions. In particular, consider again the probability-density functions \( f_k(x) \) \((k = 1, 2, \ldots, K)\) that are assumed to underlie these distributions. We
**Figure 3.** Example of a case to which the multinomial maximum likelihood mixture (MMLM) analysis applies. [The top and bottom panels, which correspond to the partitioned basis conditions of an experiment, illustrate the probability-density functions of two pure basis distributions, $f_1(x)$ and $f_2(x)$, respectively. The middle panel, which corresponds to a third condition, illustrates the probability-density function of a binary mixture distribution, $f(x)$, formed from mixing the basis distributions equally (mixing probability $\alpha = 0.5$). The small dots above each horizontal axis are representative empirical observations sampled from the distributions in each condition. The dashed vertical lines are the bin boundaries used for grouping the observations into subintervals as part of the MMLM analysis. The subscripted $\pi$ values are theoretical bin probabilities associated with each distribution; $\pi_{ik}$ denotes the probability that an observation sampled in the $k$th condition ($k = 1, 2, 3$) will fall in the $i$th bin ($i = 1, 2, \ldots, 6$).]
Table I
Frequency Matrix $N$ of Observations in the Conditions
Illustrated by Figures 3 and 4

<table>
<thead>
<tr>
<th>Bin</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Note. The cell entries in the table correspond to the frequency values $n_{ik}$ discussed in the text, where $i$ denotes the bin number ($i = 1, 2, \ldots, 6$) and $k$ denotes the condition number ($k = 1, 2, 3$).

Proceed as if each such function has a set of bin probabilities $\pi_{ik}$ ($i = 1, 2, \ldots, I$) associated with it (Figure 3). Here $\pi_{ik}$ represents the theoretical probability that an observation sampled in the $k$th condition will fall in the $i$th bin; in other words, $\pi_{ik} = F_k(x_i) - F_k(x_{i-1})$, where $F_k(x)$ is the theoretical cumulative distribution function for the $k$th condition, $x_i$ is the cutpoint for the upper boundary of the $i$th bin, and $x_0$ is the cutpoint for the lower boundary of the $i$th bin. Some examples of these bin probabilities appear in Figure 3, where the $\pi$ values in the various bin intervals illustrate the theoretical probabilities of obtaining observations (small dots) in the corresponding bins (dashed vertical lines). With this parameterization, one of our major objectives is then to estimate the bin probabilities for each basis and mixture distribution. In achieving this objective, the MMLM analysis permits us to determine the shapes of these distributions to at least a first approximation, while also providing a goodness-of-fit test for the obtained data.

Given the relations that are hypothesized to exist between the various basis and mixture distributions, certain equalities should hold among the bin probabilities for these distributions. To express the relevant equalities formally, we supplement our notation further. First, let us denote the bin probabilities for the pure basis distributions (i.e., $f_1(x), f_2(x), \ldots, f_I(x)$) associated with the $J$ empirical basis conditions as $\pi_{il}^f$ ($l = 1, 2, \ldots, I; j = 1, 2, \ldots, J$). In terms of the previous bin-probability notation, it is therefore the case that $\pi_{il}^f = \pi_{ij}$ whenever $1 \leq k = j \leq J$. Second, let us introduce an augmented set of mixing probabilities $a_{jk}$ ($j = 1, 2, \ldots, J; k = 1, 2, \ldots, K$) for the $J$ basis distributions and $K$ experimental conditions. Here $a_{jk}$ denotes the proportional contribution of the $j$th basis distribution to observations in the $k$th condition. For all values of $k$ ($1 \leq k \leq K$), $\sum_j a_{jk} = 1$ and $a_{jk} \geq 0$.

Moreover, when $k$ refers to one of the basis conditions (i.e., $1 \leq k \leq J$), the values of the mixing probabilities are constrained even further. In particular, if $1 \leq k \leq J$ and $k = j$, then $a_{jk} = 1$. This follows because, by definition, the $j$th condition in this set provides a pure sample from the $j$th basis distribution. For the same reason, if $1 \leq k \leq J$ and $k \neq j$, then $a_{jk} = 0$. For each of the $K - J$ conditions ($C_j; k = J + 1, J + 2, \ldots, K$) that involve nondegenerate mixtures of the $J$ basis distributions, it is still the case, of course, that at least two of the corresponding mixing probabilities must have values greater than zero and less than one (i.e., whenever $J + 1 \leq k \leq K$, there exist $j$ and $j'$ such that $1 \leq j \neq j' \leq J, 0 < a_{jk}, 0 < a_{jk'}$).

Consequently, we may express the original full set of theoretical bin probabilities ($\pi_{ik}^f; i = 1, 2, \ldots, I; k = 1, 2, \ldots, K$) for all $K$ experimental conditions in terms of the bin probabilities for the $J$ basis conditions ($\pi_{il}^f; i = 1, 2, \ldots, I; j = 1, 2, \ldots, J$) and the augmented set of mixing probabilities ($a_{jk}; j = 1, 2, \ldots, J; k = 1, 2, \ldots, K$). The form of the expression is simply

$$\pi_{ik} = \sum_{j=1}^{J} a_{jk} \pi_{ij}^f. \quad (16)$$

This follows directly from the definitions of a mixture distribution (Equation 1) and the respective bin probabilities.

---

Figure 4. A set of histograms obtained by grouping the hypothetical observations (small dots) above the horizontal axes in Figure 3. (The top, middle, and bottom panels illustrate the approximate shapes of the corresponding probability-density functions for the theoretical basis and mixture distributions. The subscripted $n$ values are empirical frequency counts of the observations located in each bin of each experimental condition.)

---

A degenerate mixture is formed from a single basis distribution; it is a mixture only in a formal sense.
Equation 16 constitutes the stochastic mixture model with which we will work henceforth. Using the empirical frequency matrix N of observations in the J bins for each of the K experimental conditions (e.g., Table 1 and Figure 4), our MMLM analysis is designed to yield estimates of the theoretical bin probabilities associated with the pure basis distributions (i.e., \( \pi^a \) and, equivalently, \( \pi^b \)), where \( i = 1, 2, \ldots, I; j = 1, 2, \ldots, J; 1 \leq k \leq J \). The analysis also yields estimates of the mixing probabilities (i.e., \( \alpha_k; J + 1 \leq k \leq K \)), and—through Equation 16—the bin probabilities associated with the mixture distributions formed from the basis distributions (i.e., \( \pi_k; J + 1 \leq k \leq K \)).

In applying the estimation procedure, we simply replace the theoretical parameters of Equation 16 with a set of corresponding estimators:

\[
p_{ik} = \sum_{j=1}^{J} a_{jk} \hat{p}_j^a,
\]

where the \( p_{ik}, a_{jk}, \) and \( \hat{p}_j^a \) respectively denote estimators for the mixture-distribution bin probabilities, mixing probabilities, and basis-distribution bin probabilities. Equation 17 can also be expressed more generally as a product of three matrices:

\[
P = PA A^T
\]

where \( P \), \( P^* \), and \( A \) respectively contain vectors of estimated bin probabilities (\( p_{ik} \) and \( \hat{p}_j^a \)) and mixing probabilities (\( \alpha_k \)) as shown in Table 2.

Various types of estimators could be used to instantiate Equations 17 and 18. For present purposes, the MMLM analysis relies on maximum likelihood estimators (e.g., Bishop, Fienberg, & Holland, 1975). We next discuss how maximum likelihood estimators for the theoretical bin probabilities and mixing probabilities are obtained with the MMLM analysis. The details of the following discussion may be understood more readily by referring to Table 3, which summarizes the notation introduced so far.

**Maximum Likelihood Estimation**

Our procedure for estimating the mixing probabilities and bin probabilities in Equation 17 involves an iterative hill-climbing algorithm that yields a set of maximum likelihood estimates. Because the observations that contribute to the frequency matrix N are assumed to be independent and identically distributed, each individual frequency cell \( n_{ik} \) may be treated as the realization of a binomial random variable whose parameters are \( n_k \) and \( \pi_{ik} \). Furthermore, any vector \( N_k \) may be treated as the realization of a multinomial random variable whose parameters are \( n_k \) and a vector of bin probabilities \( \pi_k = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{iJ}) \), where \( \sum_{i=1}^{J} \pi_{ik} = 1 \). By treating the data in this way, we can construct a likelihood function \( L \), and seek those values of \( p_j^* \) and \( a_{jk} \) that maximize \( L \) given the observed values of \( n_{ik} \). The obtained values are maximum likelihood estimates for the parameters in Equation 16. The mathematical details of the algorithm and its derivation are described by Smith et al. (1991). Here, we briefly sketch how the algorithm works.

**Algorithm for maximum likelihood estimation.** The MMLM algorithm alternates iteratively between estimating the theoretical bin probabilities (\( \pi_{ik} \) and \( \pi^a \)) and the mixing probabilities (\( \alpha_k \)) for the \( K \) experimental conditions. On its first pass, the algorithm estimates the bin probabilities of the basis distributions directly from the data in the \( J \) basis conditions. Next, fixing these estimates temporarily, the algorithm calculates approximate best estimates of the mixing probabilities from the data in all \( K \) conditions, given the current bin-probability estimates. After initial estimates of the mixing probabilities are obtained, a second pass is made to reestimate the basis bin probabilities from the full data set, while keeping the mixing-probability estimates temporarily fixed. Then, with the reestimated bin probabilities in hand, the mixing probabilities are estimated again. The calculations continue cycling like this until the estimated probabilities stabilize. It can be shown that the algorithm will always converge to a set of unique maximum likelihood estimates (Smith et al., 1991).

**Geometry of mixture distributions.** To understand further what the algorithm for the MMLM analysis does, it is helpful to think in terms of a geometric representation. This representation entails treating the theoretical bin probabilities \( (\pi_{ik}; i = 1, 2, \ldots, I) \) associated with each hypothesized basis and mixture distribution \( f_b(x) \) as a vector in an I-dimensional space, where \( I \) equals the number of bins created. As part of the representation, the observed bin frequencies from each condition (i.e., the values of \( n_{ik} \) in the \( k \)th column of the frequency matrix \( N \)) are converted to corresponding relative frequencies \( (q_{ik} = n_{ik}/n_k) \) and treated as vectors in the same space. Estimating the mixing probabilities and bin probabilities that best fit the data then amounts to identifying a particular subset of points on a hyperplane (i.e., a multidimensional plane) that comes closest to the data in a maximum likelihood sense.

For example, consider Figure 5. Here we have diagrammed the geometric representation of a case involving three bins (\( I = 3 \)), which contain observations having small, intermediate, and large values, respectively (e.g., RTs less than 300 ms, between 300 and 500 ms, and greater than 500 ms). These bins are represented by the three orthogonal axes in Panel A of Figure 5. Values along the axes indicate the relative frequencies with which observations are found to occur in the respective bins. Also shown in Panel A is a set of three vectors \( (Q_k; k = 1, 2, 3) \). The endpoints of these vectors correspond to empirical distributions of observed relative bin frequencies \( (q_{ik} = n_{ik}/n_k) \) for the three different experimental conditions, two of which correspond to putative basis distributions (i.e., \( Q_1 \) and \( Q_2 \)) and one of which corresponds to a putative mixture distribution (i.e., \( Q_3 \)).

More precisely, suppose that \( Q_k = (q_{ik}, q_{2k}, q_{3k}) \) for \( k = 1, 2, 3 \). Suppose further that \( n_{ik} \) is the frequency of observations in the \( i \)th bin of the \( k \)th condition, and \( n_k \) is the total number of observations in the \( k \)th condition. Then \( q_{ik} = n_{ik}/n_k \) for all \( i \) and \( k \). Consequently, as Panel A of Figure 5 shows, the vector endpoints must all fall somewhere on a triangular subplane with corners at the points \([1, 0, 0], [0, 1, 0] \), and \([0, 0, 1]\) in the three-dimensional space. This is because the relative bin frequencies of any distribution must sum to one (i.e., \( \sum_{i=1}^{I} q_{ik} = 1 \) for all \( k \)).

Now in the context of Figure 5, one may view the observed

---

7 Because \( \pi_{ik} = \pi^a_j \) whenever \( k = j \), it is also the case that \( p_{ik} = \hat{p}_j^a \) whenever \( k = j \).
Table 2
Matrices of Estimated Bin Probabilities (P and P*) and Mixing Probabilities (A) Used for Generalizing Equation 17 of the Multinomial Maximum Likelihood
Mixture Analysis (cf. Equation 18)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>P*</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I×K)</td>
<td>(p_{11} ) ( \ldots ) (p_{1k} )</td>
<td>(p_{11}^* ) ( \ldots ) (p_{1j}^* )</td>
<td>(a_{1j+1} ) ( \ldots ) (a_{1k} )</td>
</tr>
<tr>
<td></td>
<td>(p_{21} ) ( \ldots ) (p_{2k} )</td>
<td>(p_{21}^* ) ( \ldots ) (p_{2j}^* )</td>
<td>(0 ) ( \ldots ) (a_{2j+1} ) ( \ldots ) (a_{2k} )</td>
</tr>
<tr>
<td></td>
<td>(p_{31} ) ( \ldots ) (p_{3k} )</td>
<td>(p_{31}^* ) ( \ldots ) (p_{3j}^* )</td>
<td>(0 ) ( \ldots ) (0 ) ( \ldots ) (a_{3j+1} ) ( \ldots ) (a_{3k} )</td>
</tr>
<tr>
<td></td>
<td>(p_{41} ) ( \ldots ) (p_{4k} )</td>
<td>(p_{41}^* ) ( \ldots ) (p_{4j}^* )</td>
<td>(p_{41} ) ( \ldots ) (p_{4j} ) ( p_{4j+1} ) ( \ldots ) (p_{4k} )</td>
</tr>
</tbody>
</table>

relative-frequency vectors \(Q_1\) and \(Q_2\) as each manifesting a distinct underlying basis distribution, whereas the vector \(Q_3\) manifests a binary mixture distribution formed from these bases. Analogously, the theoretical bin probabilities of the basis and mixture distributions may themselves be viewed as corresponding to implicit three-dimensional vectors, \(\Pi_{11}^*\), \(\Pi_{22}^*\), and \(\Pi_{33}^*\), such that \(\Pi_{11}^* = [\pi_{11}^*, \pi_{12}^*, \pi_{13}^*]\); \(\Pi_{22}^* = [\pi_{21}^*, \pi_{22}^*, \pi_{23}^*]\); and \(\Pi_{33}^* = [\pi_{31}^*, \pi_{32}^*, \pi_{33}^*]\). Because the cell entries for each of these bin-probability vectors must sum to one, their endpoints necessarily fall on the same triangular subplane that the empirical relative-frequency vectors \(Q_k\) do. So from this perspective, what we ultimately seek are vector estimates \(p_{11}^*, p_{12}^*, \ldots, p_{1K}^*\) and \(P^*\) of the respective theoretical bin-probability vectors \(\Pi_{11}^*, \Pi_{22}^*, \Pi_{33}^*\), and \(\Pi_1\) that come closest to the relative frequencies in a maximum likelihood sense.

Panel B of Figure 5, which shows a frontal view of the triangular subplane in Panel A, illustrates geometrically how the MMLM analysis achieves this objective. In the present example, the algorithm begins by identifying a line segment between the endpoints of the two relative-frequency vectors, \(Q_1\) and \(Q_2\), that are thought to manifest the respective underlying basis distributions. This line segment is of special interest because, by definition, any point on it corresponds to some mixture of the two bases. In particular, the line segment’s endpoints provide the initial estimates of the basis bin probabilities. An estimate of the binary mixing probability \(\alpha\) is obtained by finding the intermediate point on the line segment that comes closest in a maximum likelihood sense to the endpoint of the empirical relative-frequency vector \(Q_3\), which is thought to manifest the putative mixture distribution. To be specific, the initial \(\alpha\) estimate would equal the proportion of distance along the line segment between the obtained intermediate point and the endpoint of the vector \(Q_1\) (i.e., the relative-frequency vector associated with the first basis condition).

Next, our algorithm finds another nearby line segment on the same triangular subplane such that the segment’s endpoints and one of its intermediate points, determined from the initial estimate of \(\alpha\), come even closer in a maximum likelihood sense to

Table 3
List of Symbols Used in Notation for the Multinomial Maximum Likelihood Mixture Analysis

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Limits</th>
<th>Referent</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K)</td>
<td>(K \geq 3)</td>
<td>Number of experimental conditions</td>
</tr>
<tr>
<td>(J)</td>
<td>(K &gt; J \geq 2)</td>
<td>Number of basis distributions</td>
</tr>
<tr>
<td>(I)</td>
<td>(I &gt; J)</td>
<td>Number of observation bins</td>
</tr>
<tr>
<td>(n_{ik})</td>
<td>(1 \leq i \leq l; 1 \leq k \leq K)</td>
<td>Number of observations in (i)th bin of (k)th condition</td>
</tr>
<tr>
<td>(N)</td>
<td>(I) rows, (K) columns</td>
<td>Empirical frequency matrix of (n_{ik}) values</td>
</tr>
<tr>
<td>(n_k)</td>
<td>(n_k = \sum_{i=1}^{I} n_{ik})</td>
<td>Total number of observations in (k)th condition</td>
</tr>
<tr>
<td>(f_{ij}(x))</td>
<td>(1 \leq i \leq J)</td>
<td>Theoretical basis distribution for (i)th basis condition</td>
</tr>
<tr>
<td>(f_{ij}(x))</td>
<td>(J + 1 \leq k \leq K)</td>
<td>Theoretical mixture distribution for (k)th condition</td>
</tr>
<tr>
<td>(\pi_{ik})</td>
<td>(1 \leq i \leq l; 1 \leq k \leq K)</td>
<td>Bin probability for (i)th bin in (k)th condition</td>
</tr>
<tr>
<td>(\pi_{0k})</td>
<td>(1 \leq i \leq l; 1 \leq j \leq J)</td>
<td>Bin probability for (i)th bin in (j)th basis condition</td>
</tr>
<tr>
<td>(a_{ik})</td>
<td>(1 \leq j \leq J; 1 \leq k \leq K)</td>
<td>Mixing probability for (j)th basis distribution of (k)th mixture</td>
</tr>
<tr>
<td>(\rho_{ik})</td>
<td>(1 \leq i \leq l; 1 \leq k \leq K)</td>
<td>Estimator for (\pi_{ik})</td>
</tr>
<tr>
<td>(\rho_{0k})</td>
<td>(1 \leq i \leq l; 1 \leq k \leq K)</td>
<td>Estimator for (\pi_{0k})</td>
</tr>
<tr>
<td>(\alpha_{ik})</td>
<td>(1 \leq j \leq J; 1 \leq k \leq K)</td>
<td>Estimator for (a_{ik})</td>
</tr>
<tr>
<td>(P)</td>
<td>(I) rows, (K) columns</td>
<td>Matrix of (\rho_{ik}) estimators</td>
</tr>
<tr>
<td>P*</td>
<td>(I) rows, (J) columns</td>
<td>Matrix of (\rho_{0k}) estimators</td>
</tr>
<tr>
<td>A</td>
<td>(J) rows, (K) columns</td>
<td>Matrix of (\alpha_{ik}) estimators</td>
</tr>
</tbody>
</table>

Note. For \(1 \leq j \leq J, 1 \leq k \leq K\), the following relations hold: \(\pi_{0k} = \pi_{0j}^*\), \(p_{0k} = p_{0j}^*\), \(a_{ik} = 1\), and \(a_{ik} = 1\). For \(1 \leq j \leq J, 1 \leq k \leq K\), and \(j \neq k\), the following relations hold: \(a_{ij} = 0\) and \(a_{ik} = 0\).
the empirical relative-frequency vectors \((Q_1, Q_2, \text{ and } Q_3)\). The endpoints of this new line segment are revised estimates of the theoretical bin probabilities. The process then recycles iteratively, each time adjusting its estimate of the mixing probability, finding a new line segment whose endpoints and \(\alpha\)-determined intermediate point fit the data better, and continuing until the obtained goodness-of-fit stabilizes.

As a result, a sequence of line segments and estimated \(\alpha\)-values will emerge. The endpoints of the final segment obtained before termination constitute our algorithm's best estimates of the theoretical bin probabilities for the basis distributions \((P^*\) and \(P^?)\). The intermediate point on this line defined by the final \(\alpha\)-value constitutes the best estimate of the bin probabilities for the binary mixture distribution \((P_k)\) formed from these bases. An illustration of such estimates is depicted by the dashed line and solid points in Panel B of Figure 5.

More generally, this type of geometric representation may be extended to cases involving \(I\) bins \((I > 3)\), \(J\) basis distribution \((J > 2)\), and \(K - J\) mixture distributions \((K > J)\). Although it is difficult to illustrate such cases graphically, they involve the same logic as Figure 5 does. Our algorithm iteratively seeks a simplex (i.e., a subspace analogous to the dashed line in Panel B of Figure 5) having \(J - 1\) dimensions embedded in a hyperplane (i.e., an analog of the triangular subplane in Figure 5) having \(I - 1\) dimensions such that a derived set of vector endpoints and intermediate points associated with the hypothesized theoretical basis and mixture distributions on the simplex come closest to the corresponding empirical relative-frequency vectors.

**Goodness of Fit**

To quantify the goodness of fit achieved with our algorithm, we use the likelihood-ratio statistic \(G^2\) (Bishop, Feinberg, & Holland, 1975, pp. 125ff), which is defined as

\[
G^2 = -2 \sum \text{obs} \ln \frac{\text{est}}{\text{obs}} = 2 \sum \text{obs} \ln \frac{\text{obs}}{\text{est}}
\]

\[
= 2 \sum \sum_{k=1}^{K} n_{ik} \ln \frac{n_{ik}}{n_k p_{ik}}. \tag{19}
\]

In this equation, \(\text{obs}\) represents the observed bin frequency and \(\text{est}\) represents the corresponding maximum likelihood estimate. For our application, \(n_{ik}\) is the frequency of observations in the \(i\)th bin of the \(k\)th condition, \(n_k\) is the total number of observations in the \(k\)th condition, \(p_{ik}\) is the maximum likelihood estimate of the theoretical bin probability, and \(n_{ik} p_{ik}\) is the estimated bin frequency under the present assumptions. It can be
shown that maximum likelihood estimation procedures minimize \( G^2 \) (Bishop et al., 1975).

Assuming that the data indeed come from a family of basis and mixture distributions, the statistic \( G^2 \) has an asymptotic central \( \chi^2 \) distribution whose degrees of freedom (\( df \)) are given by

\[
df = K(I - 1) - J(I - 1) - (K - J)(J - 1)
= (K - J)(I - J).
\]

The middle portion of Equation 20 includes three major components: \( K(I - 1), J(I - 1), \) and \( (K - J)(J - 1) \), where \( I, J, \) and \( K \) have been defined previously (Table 3). The first component, \( K(I - 1) \), represents the total number of condition-by-bin combinations with a priori unconstrained relative frequencies. This number comes from the fact that there are \( K \) different conditions, and within each condition, there are \( I \) relative bin frequencies that must sum to one, leaving \( I - 1 \) “free” bins in each condition. From the product of \( K \) and \( I - 1 \), two components must be subtracted, reducing the available degrees of freedom because of the parameter estimation that takes place as part of the MMLM analysis. The second major component, \( J(I - 1) \), represents the number of theoretical bin probabilities to be estimated for the basis distributions. This number comes from the fact that there are \( J \) basis distributions, and each of these distributions has \( J \) bin probabilities that must sum to one, leaving \( J - 1 \) bin probabilities per distribution to be estimated. Finally, the third component of Equation 20, \( (K - J)(J - 1) \), represents the number of mixing probabilities \( (\alpha_k) \) to be estimated. This number comes from the fact that there are \( K - J \) nondegenerate mixture distributions formed from the basis distributions, and for each mixture distribution, \( J \) mixing probabilities are required to weight the contributions of the \( J \) basis distributions; these mixing probabilities must sum to one, leaving \( J - 1 \) of them per putative mixture distribution to be estimated. When we subtract the total number of parameters to be estimated (basis bin probabilities and mixing probabilities) from the total number of condition-by-bin combinations with a priori unconstrained relative frequencies, \( (K - J)(J - 1) \) degrees of freedom remain for the \( G^2 \) goodness-of-fit test. For example, in the simple case illustrated by Figure 5, we have \( I = 3, J = 2, \) and \( K = 3, \) so \( df = (3 - 2)(3 - 2) = 1 \). This is the least complex case that still permits a nondegenerate goodness-of-fit test.\(^4\) The reason that the number of bins per distribution must exceed the number of basis distributions, as stated earlier, is now clear: if \( I \leq J, df < 1, \) and the MMLM analysis cannot be used.

To test for the presence of a mixture-distribution family, including its basis distributions and mixing probabilities, one would calculate the value of \( G^2 \) using Equation 19 and determine the degrees of freedom using Equation 20. The resulting values may then be compared with a standard table of significance levels for the \( \chi^2 \) distribution (e.g., Hays, 1981). If the comparison does not yield an outcome in the extreme upper tail of the distribution, corresponding to some prespecified significance level (e.g., \( p < .05 \)), then the hypothesis that the data come from a mixture-distribution family may be accepted at least provisionally (or, more precisely, not rejected). Otherwise, the hypothesis should be rejected.

Of course, this leads directly to a question about the statistical power of the goodness-of-fit test provided by the MMLM analysis: What is the probability that the \( G^2 \) statistic defined in Equations 19 and 20 will correctly reject the hypothesis of a mixture-distribution family when the hypothesis is false? We address this question in the next section.

### Statistical Power of the MMLM Analysis

An important outcome of the MMLM analysis is the acceptance or rejection of the null hypothesis that a mixture-distribution family adequately characterizes a set of obtained data. If the null hypothesis is rejected, then the statistical significance level chosen in advance (e.g., \( p < .05 \)) specifies the probability of a Type I error (i.e., rejecting the null hypothesis when it is true). However, when the \( G^2 \) statistic does not reach the chosen significance level, there is no way to specify directly the probability of a Type II error (i.e., accepting the null hypothesis when it is false). The latter value depends on the power of the goodness-of-fit test, which is defined as \( 1 - P(\text{Type II error}) \) (e.g., Cohen, 1988). Because \( P(\text{Type II error}) \) is a function of exactly what alternative hypothesis holds when the null hypothesis does not, it is impossible to state with complete generality what the test’s power may be.

Nevertheless, we need some indication of how much power the MMLM analysis has. There are many situations (an infinite number of them, in fact) for which the present goodness-of-fit test will prescribe accepting the null hypothesis when it is false. Such situations may arise, for instance, when the basis distributions are sufficiently similar to one another that they cannot easily be distinguished statistically. In these situations, the power of the present test, and our confidence in accepting the null hypothesis, would be low. On the other hand, a test with high power would lead to increased confidence in the inference. Thus, some power assessment is clearly required.

To help satisfy this requirement, we have conducted a series of stochastic computer simulations with the MMLM analysis and its goodness-of-fit test. For purposes of exposition, these simulations focused specifically on the power of the test to reject correctly the hypothesis of a binary mixture-distribution family (Equation 2). The following subsections briefly outline how the simulations were conducted and what results were obtained.\(^6\)

#### Stochastic simulations

Our stochastic simulations were performed in a series of trial blocks. For each block of trials, we selected three normal distributions, as illustrated in Figure 6. The three distributions had the same standard deviation \( (\sigma) \) and different means \( (\mu_1 < \mu_2 < \mu_3) \). They were selected to test a null

---

\(^1\) The degree to which the \( G^2 \) statistic's distribution approximates a central \( \chi^2 \) distribution under a valid mixture hypothesis depends on the total sample size \( (\Sigma n_a) \) and, more important, on the individual bin frequencies \( (n_a) \). For the approximation to be close, the bin frequencies should equal or exceed five observations per bin (cf. Hays, 1981). Although this rule should not be followed too rigidly, it may provide a useful guideline. Neglecting it can cause small changes in the empirical bin frequencies to yield large differences in the estimated bin or mixing probabilities.

\(^6\) See Everitt (1981) for a similar simulation study involving ways to estimate the number of bases in mixtures of normal distributions.
hypothesis that the distribution with the intermediate mean ($\mu_3$) constituted a binary mixture of the other two distributions with the extreme means ($\mu_1$ and $\mu_2$). By construction, the null hypothesis was false under these circumstances, and our objective was to assess how frequently the MMLM analysis would correctly reject this hypothesis on the basis of observations generated randomly from the three distributions.

Fifty trials were included in each block. On each trial, we generated a sample of $n_t$ observations pseudorandomly from each distribution selected for the block. The samples generated from the distributions with the smallest and largest means were treated as if they manifested two putative basis distributions. The sample generated from the distribution with the intermediate mean was treated as manifesting a putative binary mixture of the upper and lower bases. Following the procedure outlined earlier, we analyzed the obtained observations to determine whether, according to Equations 19 and 20, they were consistent with the hypothesis that they came from a binary mixture-distribution family. This test, which was conducted at the .05 level of statistical significance, yielded either a positive (accept the null hypothesis) or negative (reject it) outcome. One test was conducted per trial. We repeated the testing procedure 50 times, trial by trial, generating new independent samples of observations each time. This ultimately produced some proportion of mixture-hypothesis rejections, from which the power of our goodness-of-fit test was estimated for the combination of distribution means ($\mu_i$; $i=1, 2, 3$), standard deviation ($\sigma$), and sample size ($n_t$) used during the trial block.\footnote{The FORTRAN-77 random-number generator on a VAX 11/780 computer was used to generate the observations. The generation of each observation $x$ involved two steps: First, a standard normal deviate $z$ (i.e., a number sampled from a normal distribution with zero mean and unit variance) was generated; second, we applied a linear transformation of the form $x = z + \mu_i$ where $\mu_i$ and $\sigma$ denote respectively the mean and standard deviation of the normal distribution chosen to be sampled in Figure 6.}

From block to block, we systematically manipulated the various parameters just described to determine how they influenced the obtained power estimates. Table 4 shows the range of values associated with each parameter. For present purposes, we focus on three parameters: the amount of separation ($\Delta$) between the two putative basis distributions measured in standard deviation units ($d'$), the position of the intermediate distribution ($\theta$) relative to the basis distributions, and the sample size ($n_t$ and $I$).

We separated the mean of the lower basis distribution by a difference $\Delta$ from the mean of the upper basis distribution (i.e., $\mu_1 = \mu_2 - \Delta$). The value of $\Delta$ was selected to be a constant multiple of the distributions' standard deviation (i.e., $\Delta = d\sigma$). In essence, the multiplicative constant $d'$ may be viewed as analogous to the $d'$ sensitivity parameter of signal-detection theory (Green & Swets, 1966). Its value represents the amount by which the normalized means of the upper and lower distributions differed in standard deviation units (i.e., $d' = \Delta / \sigma$). As part of our simulations, $d'$ varied between 1.0 and 3.0 in steps of 0.5 across trial blocks.

A linear equation, $\mu_3 = \theta \mu_1 + (1 - \theta) \mu_2$, was used to set the mean of the intermediate distribution in each trial block, making it analogous to the mean of a binary mixture (cf. Equation 3). The parameter $\theta$, which may be viewed as a “pseudo-mixing probability,” varied between 0.1 and 0.9 across blocks, representing the proportional size of the difference between $\mu_2$ and $\mu_3$ compared with the difference between $\mu_1$ and $\mu_2$ (i.e., $\theta = (\mu_2 - \mu_3) / (\mu_2 - \mu_1)$). Because the intermediate and putative basis distributions were identical here except for the locations of their means, we expected the obtained power of the MMLM analysis to be an approximately symmetric function when plotted against values of $\theta$ greater and less than 0.5. The inclusion of $\theta$ values greater than 0.5 was therefore redundant with the inclusion of $\theta$ values less than 0.5. Nevertheless, we used the entire range of $\theta$ values as part of the simulation process to check its reliability and validity.

Finally, we manipulated sample size, including the number of observations ($n_t$) and number of bins ($I$) per distribution. As
stated earlier, the extent to which the goodness-of-fit statistic $G^2$ approximates a central $\chi^2$ distribution depends on sample size (cf. Footnote 8).

It must be emphasized that the intermediate distributions in the present simulations were never binary mixtures of the putative upper and lower basis distributions. The higher moments of the intermediate distributions did not satisfy the requirements of the mixture-distribution hypothesis (e.g., Equation 5 for the variance did not hold). Instead, like the upper and lower distributions, the intermediate distributions always belonged to the family of normal distributions, whereas nondegenerate binary mixtures of two normal basis distributions never belong to this family (Townsend & Ashby, 1983). Thus, if our MMLM analysis were "perfect," it should always reject the null (mixture) hypothesis under the present circumstances.

To assess how powerful the analysis really is, we combined the different possible values of $\theta$, $d'$, $I$, and $n_e$ orthogonally in our simulation design. One trial block was run (with 50 trials per block) for each possible combination of parameter values, which included seven $\theta$ levels, five $d'$ levels, four $I$ levels, and three $n_e$ levels (Table 4). This yielded a total of 420 power estimates, reflecting the joint influences of distribution separation, bin arrangement, and sample size on the power of the $G^2$ goodness-of-fit test.

**Simulation results.** A representative subset of our simulation results appears in Figure 7. Here we have plotted the estimated power of the MMLM analysis as a function of the pseudomixing probability $\theta$, which reflects the position of the intermediate distribution's mean relative to those of the putative upper and lower "basis" distributions. The power estimates come from those simulations in which we set the sampled distributions' standard deviation ($\sigma$) to 50, while varying the number of observations ($n_e$) and number of bins ($I$) per distribution. In each panel of Figure 7, the power-versus-$\theta$ functions are shown with $d'$, the normalized difference between the upper and lower distributions' means, as a parameter. The three rows of panels in the figure correspond respectively to $n_e = 40, 60,$ and 90. Panels A and C involved 6 bins per distribution, Panels B, D, and E involved 8 bins, and Panel F involved 10 bins. Together, these cases illustrate how the sample size and degree of resolution in partitioning the observations affected the proportions of simulation trials (out of 50 trials per point) for which the null binary-mixture hypothesis was correctly rejected.

The six panels of Figure 7 reveal that the shapes and absolute levels of the power-versus-$\theta$ functions are similar for the different sample sizes and numbers of bins. We found that for each sample size ($n_e$) and number of bins ($I$) per distribution, the power of the MMLM analysis generally increased with $d'$. As the putative upper and lower basis distributions diverged, more and more observations sampled from the lower tail of the lower distribution and from the upper tail of the upper distribution failed to overlap those from the intermediate distribution. This progressively increased the probability that the MMLM analysis would detect that the intermediate distribution was not a true binary mixture of the upper and lower distributions. For example, when the mean of the intermediate distribution fell midway between the means of the putative bases ($\theta = 0.5$), $d'$ values of 2.5 or greater typically yielded power estimates above 0.8.

The power of the analysis also depended strongly on $\theta$, the relative position of the intermediate-distribution mean with respect to the means of the upper and lower basis distributions. Consistent with our prior expectations about the power function's symmetry, as $\theta$ increased from 0 to 0.5, power increased gradually, and then as $\theta$ increased further from 0.5 to 1, power decreased. The overall result was an inverted-U power function. Even when the separation between the upper and lower basis distributions was relatively large ($d' > 2.0$), the power tended to be low for values of $\theta$ near zero or one. This presumably happened because $\theta$ values near zero make it difficult to discriminate the intermediate distribution from the upper basis distribution, whereas $\theta$ values near one make it difficult to discriminate the intermediate distribution from the lower basis distribution; in particular, such $\theta$ values would yield estimated bin probabilities and goodness-of-fit statistics consistent with degenerate binary mixtures (i.e., ones having mixing probabilities of 0 or 1, respectively; see Footnote 6).

Figure 7 also reveals that the number of observations and number of bins per distribution had some effect on the obtained power estimates. As $n_e$ increased or $I$ decreased, the power of the MMLM analysis increased. Surprisingly, even when the numbers of observations and bins were relatively
small (e.g., \( n_e = 40 \) and \( j = 6 \)), the MMLM analysis had fairly high power (e.g., greater than 0.75) under some representative combinations of \( d' \) and \( \theta \) (i.e., \( d' \geq 2.5; 0.4 \leq \theta \leq 0.6 \)).

**Implications for experimental design.** The results of our stochastic simulations have some important implications concerning the design of experiments involving the MMLM analysis. When an investigator uses this analysis to test for the presence of a binary mixture-distribution family, certain preconditions should be satisfied so that the test has adequate power. Procedures for collecting the required data should situate the putative basis distributions such that their means differ by at least twice their standard deviation (i.e., \( d' \geq 2 \)). Also, the putative binary mixture should have a mean approximately midway between those of the putative basis distributions; that is, the data-collection process should aim toward making the mixing probability (\( \alpha \)) equal approximately one half. Given these preconditions,

---

13 Cohen (1988, p. 56) has suggested that a power value of 0.8 should be adopted as a minimal requirement for a "good" experiment (also see Cohen, 1965, pp. 98-99, for a discussion).

14 If the different putative basis distributions have unequal standard deviations (e.g., \( \sigma_1 < \sigma_2 \)), then the \( d' \) difference between their means should be calculated in terms of the largest standard deviation (e.g., \( d' = (\mu_2 - \mu_1)/\sigma_2 \)).
the resultant power will be reasonably high (e.g., 0.8 or greater). Otherwise, it will not, unless the number of sampled observations ($n_1$) and bins ($I$) per condition are extremely large.

To achieve the necessary preconditions for high power, several more specific steps are possible, including a number of standard procedures for increasing statistical power that apply in a variety of situations. First, one can attempt to maximize the effects of experimental factors that mediate the difference between the means of the putative basis distributions. For example, in RT experiments on response preparation (Meyer et al., 1985), such effects may be amplified by encouraging subjects to use advance information provided by anticipatory priming stimuli, which reduce RTs substantially. Second, one can attempt to minimize the standard deviations of the distributions in question, for example, by encouraging subjects to maintain consistent high levels of alertness and attention. A combination of these two steps, maximizing mean differences and minimizing standard deviations, will have the net result of increasing the relative separation ($d'$) between two or more putative basis distributions and so increase the power attained by the MMLM analysis.

In addition, one can attempt to position the putative intermediate mixture distributions so that they fall midway between the basis distributions from which they are thought to come. Assuming that there is an experimental factor whose levels govern the value of the underlying mixing probability ($\alpha$), this factor may be adjusted adaptively to obtain any desired value of the mixing probability (e.g., $\alpha = 0.5$). For example, in RT experiments on response preparation (Meyer et al., 1985), it has been found that the time interval between the onsets of a priming stimulus and a test stimulus determines the degree of preparation induced by the priming stimulus. An experimenter may, as a result, judiciously control the mixing probability associated with different preparation states through staircase-tracking adjustments in the length of this interval (Levitt, 1971).

Extensions to higher order mixtures. Although the present stochastic simulations do not deal directly with higher order mixture distributions, our conclusions about the power of the MMLM analysis may be generalized to cases involving three or more basis distributions. Consider, for example, a case in which it is hypothesized that there are three basis distributions and two mixture distributions formed from them. Also suppose that the mean of each putative basis distribution is at least two $d'$ units from the means of the other two basis distributions. Then we would expect the MMLM analysis to be reasonably powerful when the mean of one putative mixture distribution falls midway between the lower and intermediate basis distributions, while the mean of the other putative mixture distribution falls midway between the intermediate and upper basis distributions. The exact power of the test will, of course, depend on the precise values of the mixing probabilities and the $d'$ separations between the bases.

Analyses Without Identification of Basis Distributions

Another matter that requires further consideration concerns the partitioning of experimental conditions associated with the individual basis and putative mixture distributions in the MMLM analysis. Thus far, we have assumed that the analysis involves a total of $K$ experimental conditions, the first $J$ of which are assumed to yield observations from individual pure basis distributions ($J < K$), and the remaining $K - J$ of which are assumed to yield observations from putative mixture distributions formed from the $J$ bases. In practice, however, these assumptions may not always be completely realistic. For example, one may suspect that some obtained observations come from a set of mixture distributions, but one may not know which subsets, if any, come from pure basis distributions. Under such circumstances, a question then arises about whether the MMLM analysis would still apply.

The answer to this question is somewhat complex. If samples of observations are available from $K$ conditions, and $J$ basis distributions are thought to underlie them, but a pure sample of each basis distribution is not available, then tests of the mixture hypothesis can still be performed using the MMLM analysis. For these tests, one would (more or less arbitrarily) designate $J$ of the $K$ conditions as being associated with the individual putative bases, and one would treat the samples of observations from them as surrogate basis-distribution estimates. (The choice of which conditions to so designate may rest at least in part on ancillary theoretical grounds.) Next, one would perform the standard MMLM analysis on the derived data structures, as outlined earlier. This would yield a $G^2$ statistic based on maximizing the likelihood of the data given both the mixture hypothesis and the chosen designation of the basis and mixture conditions. Depending on whether the obtained $G^2$ statistic is significant, one could then tentatively accept or reject the hypothesis. However, one could not uniquely estimate the true theoretical probabilities or mixing probabilities for the individual basis and mixture conditions. Without accurate prior identification of the conditions, a wide range of alternative parameter estimates would yield the same ultimate goodness-of-fit for Equation 1.

In searching for the best fit, one could tentatively designate various other subsets of $J$ conditions as manifesting the individual basis distributions. For each such designation, the MMLM analysis will yield new maximum likelihood estimates for the data and an additional value of the $G^2$ statistic. Some of these may provide better fits to the mixture hypothesis than others do, depending on exactly how the conditions are partitioned. To determine what the best fit is, one would have to check systematically all possible partitions of the $K$ conditions into subsets of $J$ basis conditions and $K-J$ mixture conditions, determining which partition yields the best fit. Of course, in this testing process, the chosen statistical-significance level would also have to be adjusted appropriately to maintain adequate power.

Relation to EM Algorithm

The sort of ambiguity just described, which sometimes arises during applications of the MMLM analysis, may similarly be encountered during applications of a related statistical technique, the so-called EM algorithm (Dempster, Laird, & Rubin, 1977; Titterington et al., 1985, §4.3.2). This algorithm, whose name comes from a combination of "expectation" and "maxi-

---

15 We thank Steve Link for mentioning to us the relevance of the EM algorithm.
mization,” provides a general-purpose way of obtaining precise, though not necessarily unique, maximum likelihood parameter estimates for sets of “incomplete” data. As in the MMLM analysis, the objectives of the EM algorithm are achieved through an iterative process with multiple steps: (a) A subset of data is used to estimate some tentative parameter values; (b) from these values and the rest of the data, maximum likelihood estimates of other relevant parameters are derived; (c) the values of the parameters originally estimated in the first step are reestimated on the basis of results from the second step; and (d) the second and third steps are repeated cyclically until they stabilize, yielding a best fit between all of the parameters and data. In fact, because of the EM algorithm’s generality, one could use it instead of the MMLM analysis for estimating the theoretical bin probabilities and mixing probabilities of a mixture-distribution family.

However, the MMLM analysis has several advantages over the EM algorithm in the present context. Our approach is designed specifically to deal with the problem of analyzing mixture distributions when separate samples from each of the underlying basis distributions are available. Consequently, the iterative process of the MMLM analysis can produce parameter estimates and goodness-of-fit statistics more quickly and efficiently in the targeted realm than does the more general EM algorithm. As part of the MMLM analysis, theorems about the convergence and uniqueness of its solutions have been proven (Smith et al., 1991). Such theorems do not yet exist for corresponding applications of the EM algorithm for analyzing mixture-distribution families. It is currently unclear whether the convergence and uniqueness properties associated with such applications would make the EM algorithm an attractive alternative to the MMLM analysis, given our present objectives.

Applications of the MMLM Analysis

Because the MMLM analysis is a new approach to testing hypotheses concerning the presence of mixture distributions, it has not yet been used extensively in psychological and other applications. There are, however, a few cases that illustrate how the analysis can help select among theoretical models under which mixture-distribution families play a major role. We briefly discuss two such examples here. One of them concerns the dynamics of response preparation (Meyer et al., 1984, 1985). The other concerns the dynamics of spreading activation in semantic and episodic memory (Yantis & Meyer, 1988).

Dynamics of Response Preparation

Extending the work of previous investigators, Meyer et al. (1984, 1985) applied the MMLM analysis to assess discrete versus continuous models of response preparation. They focused on a discrete all-or-none preparation model (cf. Falmagne, 1965; Ollman, 1966; Theios & Smith, 1972; Yellott, 1967, 1971). According to this model, there are only two distinct states of preparation: fully prepared and unprepared. According to continuous models, on the other hand, response preparation increases gradually over time from an initial base level to a final asymptotic level, passing through a range of intermediate states. These alternative characterizations of the preparation process may be evaluated by determining whether or not subjects’ RT distributions embody a binary mixture of two preparatory states (i.e., unprepared and fully prepared) when an intermediate amount of time is permitted for response preparation to take place. Given intermediate preparation times, the discrete all-or-none model predicts that a binary mixture RT distribution should emerge, whereas a continuous model does not.

Meyer et al. (1984, 1985) tested these predictions by having subjects perform a two-choice reaction task in which there were various amounts of time to prepare the required responses. The subject’s task was to press a left or right finger key, respectively, when a test stimulus consisting of either a left arrow (→) or a right arrow (←) appeared on a display screen. Before each test stimulus appeared, a priming stimulus that consisted of either an English word or a pronounceable nonword was presented. The lexical status of the priming stimulus (word vs. nonword) and the identity of the immediately subsequent test stimulus (right arrow vs. left arrow) were perfectly correlated. If the priming stimulus was a word, then the test stimulus was always a right arrow, requiring a right keypress; if the priming stimulus was a nonword, then the test stimulus was always a left arrow, requiring a left keypress. Subjects were encouraged, through instructions and monetary incentives, to use this contingency as much as possible for preparing the appropriate response in advance.

The time interval between the onsets of the priming and test stimuli (stimulus-onset asynchrony or SOA) was varied systematically across three priming conditions, which we call unprimed, partially primed, and completely primed. In the unprimed condition, the SOA was 0 ms, that is, no informative prime was presented; here subjects could not prepare a response in advance. Consequently, RB on every trial were presumably generated from an unprepared state.

In the completed primed condition, the SOA was 700 ms. Here, subjects always had enough time to determine the lexical status of the priming stimulus and to prepare the appropriate

---

14 The term “incomplete data” is defined as follows (Dempster et al., 1977). Suppose there are two sample spaces, \( S \) and \( S' \), with a many-to-one mapping from \( S \) to \( S' \). Also, suppose that a vector of observations \( x \) is sampled in \( S \), but this vector is only manifested indirectly as a corresponding vector of observations \( y \) sampled in \( S' \). Then the \( y \) observations are said to be incomplete data. The EM algorithm assumes that \( x \) is characterized by a family of sampling densities \( f(x | \phi) \) with parameters \( \phi \), and \( y \) is characterized by a corresponding family of densities \( g(y | \phi) \), where \( g(y | \phi) = \int f(x | \phi) dx \), the range of integration spanning the region of \( S' \), from which \( y \) may have come. Given this characterization, the EM algorithm seeks estimates of the parameters \( \phi \) that maximize \( g(y | \phi) \) for the incomplete data \( y \), taking into account the underlying densities \( f(x | \phi) \) that generated the data. Such an approach is relevant to our present concerns in that observations sampled from a mixture distribution are a type of incomplete data \( y \). These observations correspond implicitly to ones from underlying basis distributions like the \( f(x | \phi) \) that produces the \( x \) assumed by the EM algorithm. The associated bin probabilities are like the \( \phi \) parameters. As with other incomplete data, we cannot tell exactly which bases produced which mixture observations; thus, the basis bin probabilities must be estimated indirectly through the mixture distributions, analogical to the \( g(y | \phi) \).
response before the test stimulus appeared. Thus the RTs on every trial in this condition were presumably generated from a fully prepared state.

In the partially primed condition, the SOA had an intermediate duration. Discrete all-or-none and continuous models of response preparation make different predictions about what form of RT distribution should emerge under this condition. According to a discrete all-or-none preparation model, as mentioned earlier, only two preparatory states are possible: fully prepared and unprepared. The transition between these states is assumed to occur as a single discrete step at some randomly distributed point in time after the information from a priming stimulus reaches the response-preparation process. So this model predicts that the distribution of RTs under the partially primed condition with an intermediate SOA should be a binary mixture of the distributions associated with the unprimed and completely primed conditions, respectively; the unprimed and completely primed conditions would, in essence, provide distributions that are the bases for the predicted binary mixture. If the prediction fails to hold, then the discrete all-or-none preparation model must be rejected. With a continuous model, however, one would not expect the partially primed condition to yield a binary mixture distribution of RTs (see Meyer et al., 1985, for details). Thus, if the mixture prediction fails to hold, there may indeed be a gradual growth of preparation over time.

To test for the presence of binary mixture RT distributions, Meyer et al. (1985) applied the MMLM analysis to data from four subjects. In all cases, the data were consistent with the mixture hypothesis. For example, Figure 8 shows data from one representative subject when the priming stimuli were nonwords and the test stimuli were left arrows. The top, middle, and bottom panels of the figure contain relative-frequency distributions of RTs for the unprimed, partially primed, and completely primed conditions, respectively. The solid histograms represent observed RTs, and the dashed lines represent the best-fitting estimates of the basis and mixture distributions obtained through the MMLM analysis.

Several aspects of these results should be noted. First, the mean of the completely primed distribution is significantly smaller than that of the unprimed distribution (mean difference over all subjects = 75 ms). Second, the mean of the partially primed distribution is approximately midway between the means of the unprimed and completely primed distributions. Third, the variance of the partially primed distribution is larger than those of the unprimed and completely primed distributions. Fourth, the lower tail of the partially primed distribution overlaps that of the completely primed distribution, and the upper tail of the partially primed distribution overlaps that of the unprimed distribution. The estimated mixing probability (α) equalled 0.64. This pattern is exactly what one would expect if the partially primed distribution involves a binary mixture of the distributions in the unprimed and completely primed conditions (see, for example, Equations 3 and 5 and Figures 3 and 4). The goodness-of-fit test from the MMLM analysis yielded a $G^2$ value of 16.5 with 13 degrees of freedom ($p > .20$), further supporting the binary mixture hypothesis and the discrete all-or-none model of response preparation. With few exceptions, each of the other subjects exhibited the same pattern of results (see Meyer et al., 1985, for details).

---

Figure 8. Data from Meyer, Yantis, Osman, and Smith (1985, Experiment 1) illustrating the fit for a discrete two-state response-preparation model. (Subject BC produced these results when the priming stimuli were nonwords and the test stimuli were arrows pointing to the left. Top, middle, and bottom panels depict results from the unprimed, partially primed, and completely primed conditions, respectively. Solid histograms represent observed relative frequencies, and dashed lines represent the best fit of a binary mixture-distribution family [Equation 2]. The estimated mixing probability α equals 0.64. These data are consistent with the discrete two-state model [$G^2(13) = 16.5$, $p > .20$]. From "Temporal Properties of Human Information Processing: Tests of Discrete Versus Continuous Models," by D. E. Meyer, S. Yantis, A. Osman, & J. E. K. Smith, 1985. Cognitive Psychology, 17, p. 477. Copyright 1985 by Academic Press. Adapted by permission.)

---

17 An adaptive staircase-tracking procedure (Levitt, 1971) was used to set the duration of the priming interval on each trial. See Meyer, Yantis, Osman, and Smith (1985) for details.

18 The power of these tests was reasonably high. Each basis distribution in the experiment of Meyer, Yantis, Osman, and Smith (1985) had a standard deviation of about 30 ms, whereas the difference between the means of the basis distributions averaged 75 ms, yielding a t-value of 2.5. Furthermore, the adaptive staircase-tracking procedure used
Dynamics of Spreading Activation in Semantic Memory

Complementing the research by Meyer et al. (1984, 1985) on response preparation, Yantis and Meyer (1988) have applied the MMLM analysis to study the dynamics of spreading activation in semantic memory. Their experiments focused on whether the growth of activation during the retrieval of stored lexical information is discrete or continuous. In these experiments, subjects made lexical (word–nonword) decisions about strings of letters that formed words on half the trials and pronounceable nonwords on the other half (cf. Meyer & Schvaneveldt, 1971, 1976). At various moments before the onset of each test stimulus, a priming word (e.g., doctor) was presented. When the test stimulus was also a word, it had a meaning highly related to the priming word (e.g., nurse presented after doctor).

As in Meyer et al. (1985), there were three priming conditions: unprimed, partially primed, and completely primed, which involved varying the SOA between the priming and test stimuli. Yantis and Meyer (1988) tested two alternative classes of models for activation dynamics. One class involved discrete all-or-none models in which spreading activation is assumed to increase from an initial base level to a final asymptotic level in a single step at each relevant node of semantic memory (e.g., see Anderson, 1976; Collins & Quillian, 1969). These models assume that the moment of transition varies randomly from trial to trial, and that the memory node of a test stimulus may be in only one of two possible activation states on each trial. Accordingly, the lexical-decision RTs under the partially primed condition should be a binary mixture of the distributions from the unprimed and completely primed conditions, reflecting contributions from the two possible activation states.

In contrast, a second class of models tested by Yantis and Meyer (1988) assumes that spreading activation increases gradually over time at each node of semantic memory, passing through a continuum of intermediate levels (e.g., see Anderson, 1983; McClelland, 1979; McClelland & Rumelhart, 1981; Meyer & Schvaneveldt, 1971, 1976; Ratcliff, 1978). Continuous models, unlike the discrete all-or-none models, predict that a binary-mixture distribution of RTs should not emerge under the partially primed condition. Instead, this condition should yield RTs whose mean, variance, and tails all fall between those observed under the unprimed and completely primed conditions.

Yantis and Meyer (1988) found that the mixture hypothesis was soundly rejected by data from 6 subjects; none of their RT distributions were consistent with a discrete all-or-none model of activation. For example, Figure 9 shows the results from 1 representative subject. Unlike the data shown in Figure 8 (Meyer et al., 1985), the upper and lower tails of the partially primed distribution do not overlap completely with the corresponding tails of the unprimed and completely primed distributions, and the variance of the hypothesized mixture distribution is not larger than those of the putative basis distributions. The dashed lines in Figure 9, which represent the best fitting family of mixture distributions, deviate markedly from the observed frequency data (Figure 9, solid bars). As a result, the MMLM analysis yielded a significantly poor fit \( G^2(8) = 48.8, p < .001 \).

Summary

The results of Yantis and Meyer (1988) provide an interesting theoretical contrast to those of Meyer et al. (1984, 1985). Whereas response preparation appears to be a discrete all-or-
none process in an elementary choice-reaction task, the spread of activation in semantic and episodic memory may entail a more gradual, continuous process. Taken together, the studies by Meyer et al. (1984, 1985) and by Yantis and Meyer (1988) illustrate the potential value of the MMLM analysis in testing alternative psychological models.

Conclusion

The MMLM analysis offers a useful complement to previous approaches for testing hypotheses involving mixture distributions and the stochastic models underlying them. Several virtues of the analysis are noteworthy. First, unlike many previous approaches, it makes no strong assumptions about the functional forms of putative basis distributions and their mixtures. Second, it may be used in situations involving not only binary mixtures but also higher order mixtures. Third, it takes into account the entire shapes of these distributions, and yields relatively precise estimates of them along with associated mixing probabilities. Fourth, quantitative goodness-of-fit statistics, not just informal qualitative assessments, may be obtained from the analysis. Fifth, although the analysis requires sampling some data from pure basis distributions to be most effective, it may still apply across a wide variety of situations. Sixth, under appropriate circumstances, the analysis is reasonably powerful in discriminating valid versus invalid mixture-distribution hypotheses. Given these virtues, the MMLM analysis may prove useful in a variety of psychological research, including studies of human cognition and action.

References


Received August 18, 1989
Revision received January 17, 1991
Accepted January 21, 1991