

A likelihood ratio test for mixture effects

JEFF MILLER

University of Otago, Dunedin, New Zealand

Under certain circumstances, it is theoretically important to decide whether a difference between two conditions in mean reaction time (RT) results from a relatively uniform slowing of all the responses in the slower condition or from a mixture of some slowed trials with some unslowed ones. This article describes a likelihood ratio test that can be used to differentiate between these two possibilities and reports computer simulations examining the power and Type I error rate of the test under conditions similar to those encountered in RT research. A freely available computer program, called MIXTEST, can be used both to carry out the likelihood ratio test and to conduct simulations evaluating the performance of the test within various settings.

In many areas of reaction time (RT) research, probability distributions of RT have been analyzed to obtain information beyond that obtainable from mean RTs alone (see, e.g., Fazio, 1990; Luce, 1986; Meyer, Osman, Irwin, & Yantis, 1988; Miller, 1982; Rouder & Speckman, 2004; Van Zandt, 2000; Wenger & Townsend, 2000). Because of the extra information available from the analysis of RT distributions, researchers have become increasingly interested in examining these distributions for clues as to the precise causes of between-condition differences in mean RT (e.g., Gottlob, 2004; Heathcote, Popiel, & Mewhort, 1991; Hockley, 1984; Hohle & Gholson, 1968; Juhel, 1993; Leth-Steensen, Elbaz, & Douglas, 2000; Mewhort, Braun, & Heathcote, 1992; Schwarz, 2001). For example, some researchers have fit the ex-Gaussian RT model (e.g., Hohle, 1965) to the RT distributions obtained in different conditions, aided in part by special software designed for this purpose (e.g., Dawson, 1988; Heathcote, 1996). Within this model, each RT is conceived of as a sum of independent Gaussian and exponentially distributed components. By fitting the model to each condition separately, it is possible to estimate the experimental effect on each of the two components separately.

The present article contributes to the trend of RT distribution analysis by addressing the question of whether an experimental effect on mean RT is a uniform effect or a mixture effect. Figure 1 illustrates the basic scenario underlying this question. Suppose that an observed mean RT is 50 msec greater in an experimental condition than in a control condition, as is illustrated in the top panel. One possibility is that this effect reflects a slowing of responses in the experimental condition that is reasonably *uniform*

across the whole distribution of RTs, as is illustrated in the left panels. Somewhat crudely, this possibility can be summarized by saying that the effect produces a shift of the whole RT distribution in the experimental condition. Alternatively, as is illustrated in the right panels, the 50-msec difference in means may reflect a *mixture* effect. As is illustrated in panel C, 75% of the trials in the experimental condition could come from the same distribution as that in the control condition; these trials could be described as being unaffected by the experimental manipulation. In addition, however, the experimental condition could include 25% of the trials from a distribution with a much larger mean; these are the effect-present trials. Naturally, the effect must be larger than 50 msec in the effect-present trials to produce an effect of 50 msec for the mixture distribution as a whole (i.e., on average across slowed and unslowed trials).

In practical experimental situations, researchers are sometimes quite interested in knowing whether an effect on mean RT is a uniform effect or a mixture effect. Two substantive examples can be mentioned briefly to illustrate this interest. One concerns the flanker compatibility effect, which arises in a focused attention task requiring participants to respond to a center target letter while ignoring irrelevant flankers on both sides of it (see, e.g., Eriksen & Eriksen, 1974). It is well established that responses are slower in an experimental condition in which the irrelevant flankers are letters assigned to a response competing with the one required by the target, as compared with a control condition in which the irrelevant flankers are not assigned to any response at all. Evidently, the irrelevant flankers activate the response to which they are assigned, and the consequent response competition slows responses to the targets. But is response competition present in every trial (i.e., is it a uniform effect), or does it instead arise only in occasional trials, perhaps due to an occasional failure to maintain attentional focus (i.e., is it a mixture effect)? The implications of the flanker effect for our understanding of attentional selection depend partly on whether it is a uniform effect or a mixture effect.

This research was supported by a grant from the Marsden Fund, administered by the Royal Society of New Zealand. I thank Ann Reynolds and Wolfgang Schwarz for constructive comments on earlier versions of the article, the software, and the software documentation. Correspondence concerning this article should be addressed to J. Miller, Department of Psychology, University of Otago, Dunedin, New Zealand (e-mail: miller@psy.otago.ac.nz).

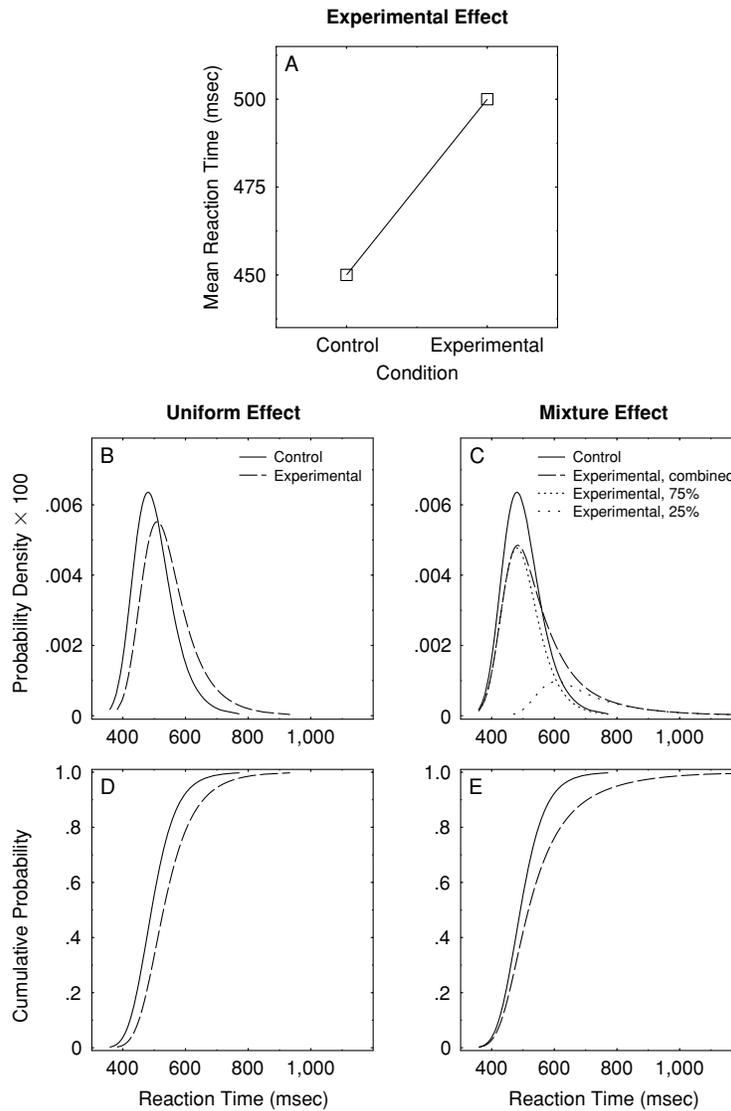


Figure 1. Illustration of two possible differences in reaction time (RT) distributions that might underlie a 50-msec effect on mean RT. The panel in the top row depicts the ambiguous 50-msec difference in mean RTs between a control condition and an experimental condition. The panels in the middle and bottom rows illustrate two different pairs of probability distributions that would produce the same 50-msec effect on mean RT, with probability densities shown in the middle row and cumulative probabilities shown in the bottom row. The left panels (B and D) illustrate control and experimental probability distributions producing the 50-msec effect in a uniform fashion, with the RT distribution in the experimental condition mainly shifted relative to that of the control condition. The right panels (C and E) illustrate probability distributions producing the 50-msec effect as a mixture, with only 25% of the trials affected by the experimental manipulation. In panel C, the probability densities of the 75% unaffected trials and 25% effect-present trials are shown as dotted lines, and these densities are scaled by the probability of each kind of trial. The “Experimental, combined” density shows the overall probability density for the experimental condition, taking into account both kinds of trials included in the mixture. The distribution in the control condition is ex-Wald with $\mu_w = 450$, $\sigma_w = 45$, and $\mu_e = 50$ for all four panels. For the uniform effect, the distribution in the experimental condition is ex-Wald with $\mu_w = 475$, $\sigma_w = 45$, and $\mu_e = 75$, so the 50-msec effect on the mean is split equally between the Wald and the exponential components of the distribution. For the mixture effect, the distribution in the experimental condition is made up of 75% trials from the control condition and 25% trials from the ex-Wald with $\mu_w = 550$, $\sigma_w = 45$, and $\mu_e = 150$.

A second example arises in the literature on the psychological refractory period (PRP) paradigm (e.g., Pashler, 1994). In this paradigm, participants are asked to perform two different tasks in each trial, and the stimulus presentations for the two tasks are separated by a small stimulus onset asynchrony (SOA). According to bottleneck models of the PRP paradigm (e.g., Welford, 1959), RT to the first task should be independent of SOA. In actual data, however, mean first-task RTs sometimes increase or decrease with SOA (e.g., Allen, Smith, Vires-Collins, & Sperry, 1998; Pashler, 1991; M. C. Smith, 1969; Tombu & Jolicœur, 2002). Proponents of bottleneck models sometimes explain these unexpected effects in terms of response grouping (e.g., Pashler & Johnston, 1989). Because it seems more convenient to emit the responses for both tasks at the same time, participants may occasionally delay the first-task response until they are also ready to respond to the second task (see, e.g., Borger, 1963). Such an explanation explicitly attributes the effect of SOA on first-task RT to a mixture effect that is present only in the occasional trials with grouped responses.

The purpose of this article is to examine formally the problem of distinguishing between uniform and mixture effects by using RT distributions. As will be considered further in the General Discussion section, previous researchers have addressed this problem only with informal examination of the probability density functions (PDFs) and cumulative probability density functions (CDFs) of observed RT distributions. The present approach is an attempt to develop and evaluate a likelihood ratio test for comparing uniform versus mixture effects.

A LIKELIHOOD RATIO TEST

In brief, likelihood ratio tests compare hypotheses in terms of the relative likelihood of the observed data under each of two hypotheses. As a trivial example, the probability of observing nine heads in 10 flips of a coin is .39 if the coin's true probability of a head is .9, whereas it is only .01 if the coin is fair. Given that the probability of the observed data (nine heads) is much higher under the hypothesis of a biased coin than under the hypothesis of an unbiased one, the likelihood ratio test would tend to support the hypothesis of a biased coin.

Likelihood ratio tests have been studied in detail in the statistical literature (e.g., Chernoff, 1954; Wilks, 1938), and they have been widely applied as a general statistical tool (e.g., Hoel, 1962; Rao, 1965). They have also been considered frequently in cognitive psychology, including RT research (e.g., Dixon, 2003; Meyer, Yantis, Osman, & Smith, 1985). Likelihood ratio tests have even been developed for the problem of determining whether a single sample comes from a homogeneous or a mixture distribution (e.g., Chen & Chen, 2001), a problem that is more difficult than the current one, because of the lack of a fixed control distribution.

In the most general case for the present problem, the control and experimental condition RTs have distinct dis-

tributions with PDFs $f_c(t)$ and $f_e(t)$, respectively. The likelihood of a set of N_c control condition RTs, $\{X_1, X_2, \dots, X_{N_c}\}$, is thus

$$\Gamma_c = \prod_{i=1}^{N_c} f_c(X_i). \quad (1)$$

Similarly, the likelihood of a set of N_e experimental condition RTs, $\{Y_1, Y_2, \dots, Y_{N_e}\}$, is

$$\Gamma_e = \prod_{i=1}^{N_e} f_e(Y_i). \quad (2)$$

Given the independence of the observations in these two conditions, the likelihood of the full data set is simply the product of these two—namely, $\Gamma_c \cdot \Gamma_e$.

According to the hypothesis of a uniform effect, each of f_c and f_e can, in principle, be any arbitrary probability distribution—denote these as f_{cu} and f_{eu} , respectively. In practical RT research, these would generally be expected to have similar forms—for example, both ex-Gaussian—although their parameters would certainly differ. Thus, the likelihood of the data under this hypothesis is simply

$$\Gamma_u = \prod_{i=1}^{N_c} f_{cu}(X_i) \times \prod_{i=1}^{N_e} f_{eu}(Y_i). \quad (3)$$

In contrast, the data have a somewhat different representation according to the alternative hypothesis of a mixture effect. Under this hypothesis, the RTs in the control condition again come from some arbitrary distribution, f_{cm} . RTs in the experimental condition, however, come from a mixture distribution. With a probability of P , the effect in question is present on an experimental trial, so the RT comes from an effect-present distribution with PDF f_{em} . On the other hand, with a probability of $1 - P$, no effect is present, so the RT comes from the control distribution with PDF f_{cm} . Thus, the PDF of experimental condition RTs under this hypothesis is given by the mixture distribution:

$$\begin{aligned} f_e(t) &= \begin{cases} f_{em}(t) & \text{with probability } P \\ f_{cm}(t) & \text{with probability } 1 - P \end{cases} \\ &= Pf_{em}(t) + (1 - P)f_{cm}(t). \end{aligned} \quad (4)$$

Under the mixture hypothesis, then, the likelihood of the whole data set is

$$\begin{aligned} \Gamma_m &= \prod_{i=1}^{N_c} f_{cm}(X_i) \\ &\times \prod_{i=1}^{N_e} [Pf_{em}(Y_i) + (1 - P)f_{cm}(Y_i)]. \end{aligned} \quad (5)$$

On the basis of the assumptions underlying Equations 3 and 5, a likelihood ratio test can be used to decide between the uniform and the mixture models for the effect in the experimental condition. Specific functional forms (e.g., ex-Gaussian) of RT distribution must be assumed for f_{cu} , f_{cm} , f_{eu} , and f_{em} . The two functional forms for the control condition (i.e., f_{cu} and f_{cm}) must be the same for both mod-

els (e.g., both ex-Gaussian), as must the two functional forms for the experimental condition (i.e., f_{eu} and f_{em}), although the functional forms for the experimental and the control conditions could, in principle, be different from one another.¹ Parameters for the two models can then be chosen to maximize the likelihood functions, Γ_u and Γ_m , and then the quantity

$$\lambda = -2 \cdot \ln \frac{\Gamma_u}{\Gamma_m} \tag{6}$$

should have an approximately χ^2 distribution (Rao, 1965; Wilks, 1938), subject to certain regularity conditions. This χ^2 distribution has one degree of freedom corresponding to the extra parameter of the mixture model (i.e., P). Thus, the null hypothesis is that the effect is uniform, and a significant observed λ would lead the experimenter to reject this null hypothesis and conclude that there was actually a mixture effect.

Computational Example

It may be useful to illustrate these concepts with a simple computational example. Table 1 exhibits a sample data set with 20 observations in both the control and the experimental conditions. To simplify the computations and emphasize that this likelihood ratio test is not limited to RT data, I have used sample scores in the range of 10–100, which might be interpreted as, for example, percentage scores on a test. Also, to keep the example as simple as possible, I have assumed that f_{cu} , f_{eu} , f_{cm} , and f_{em} are all normal distributions, so that each may be characterized fully by its mean and standard deviation (i.e., μ_{cu} , σ_{cu} , μ_{eu} , σ_{eu} , etc.).

Table 2 displays estimated parameter values obtained from the data in Table 1 via the program MIXTEST, which will be introduced at the end of this article. Note first that the estimated distribution means for the uniform model, $\hat{\mu}_{cu}$ and $\hat{\mu}_{eu}$, are simply the means of the observations in the control and the experimental conditions, respectively. This is to be expected, because a sample mean is the maximum likelihood estimate of the true mean for a normal distribution. More generally, this illustrates the fact that estimates of f_{cu} 's parameters are influenced only by observations in the control condition and estimates of f_{eu} 's parameters are influenced only by observations in the experimental condition; this fact can also be ascertained from Equation 3. Similarly, $\hat{\sigma}_{cu}$ and $\hat{\sigma}_{eu}$ are the usual maximum likelihood estimates of σ based on the data from within each condition.

Table 1
Example Data Used in Illustrating the Likelihood Ratio Test

Condition	Observed Data Values
Control	34 40 41 45 47 48 48 48 48 49 51 51 52 52 52 53 55 55 55 67
Experimental	36 38 42 43 47 49 49 54 65 74 77 83 87 87 88 92 92 93 94 95

Note—The numerical scores in the control and the experimental conditions have been sorted for ease of inspection.

Table 2
Estimated Parameter Values for the Example Data in Table 1

Parameter	Estimated Value
Uniform Model	
μ_{cu}	49.550
μ_{eu}	69.250
σ_{cu}	6.599
σ_{eu}	21.457
Mixture Model	
μ_{cm}	48.729
μ_{em}	87.322
σ_{cm}	7.258
σ_{em}	6.837
P	.553

Different estimates are obtained for the mixture model, because in this model each estimate is based on the data from both conditions. In particular, Equation 5 makes it clear that the parameters of f_{cm} should depend somewhat on the observations in the experimental condition (Y values), at least when the estimate of P differs from one.

Using the parameter estimates shown in Table 2, tables of the normal PDF, and Equations 3 and 5, it is straightforward to compute that for this example,

$$\begin{aligned} \Gamma_u &= 2.1363e - 68, \\ \Gamma_m &= 2.3891e - 65, \\ \lambda &= 14.039. \end{aligned}$$

The observed value of λ is highly significant, relative to the expected $\chi^2(1)$ distribution, so the hypothesis of a uniform effect can be rejected. The researcher can conclude that the effect reflects a mixture.

SIMULATIONS

Computer simulations were carried out in order to examine the statistical properties of the proposed likelihood ratio test statistic λ given by Equation 6 under conditions typical of those found in RT experiments. These simulations used the ex-Wald distribution as the basic underlying model for an RT distribution, as will be described in the next section. The simulation conditions and procedure will be described in the subsequent two sections.

The Ex-Wald Distribution

Schwarz (2001) proposed the ex-Wald distribution as a theoretically attractive and practically useful model for RT distributions. In brief, each sample from an ex-Wald distribution can be conceived of as the sum of two independent random variables, one from the Wald distribution and one from the exponential distribution.² Thus, the ex-Wald is quite similar to the better known ex-Gaussian, with a Wald component replacing the Gaussian one.

The ex-Wald distribution has attracted interest (e.g., Heathcote, 2004) because it has theoretical advantages, relative to the more traditional ex-Gaussian distribution. Its biggest theoretical advantage is that the Wald component has a natural interpretation as a distribution of laten-

cies, whereas the Gaussian—which can even be negative, in theory—does not. Specifically, the Wald distribution (also known as the inverse Gaussian distribution) is the distribution of first passage times for a Wiener diffusion process to reach a given criterion cutoff value (cf. Chhikara & Folks, 1989; Cox & Miller, 1965). Models based on diffusion processes have had considerable success in accounting for basic phenomena in RT and response choice (e.g., Diederich, 1995; Ratcliff, 1978; Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Rouder, 1998; Schwarz, 1993, 1994; P. L. Smith, 2000; Usher, Olami, & McClelland, 2002), so the Wald seems much more plausible as a component of RT than does the Gaussian. In addition, a secondary theoretical advantage of the ex-Wald is that changes in the Wald component can, in principle, be partitioned among effects on drift rate versus criterion value (see Schwarz, 2001), which is in many ways analogous to partitioning sensitivity versus decision effects within signal detection theory (see, e.g., Green & Swets, 1966).

In practical terms, another attractive feature of the ex-Wald distribution is that it is quite similar to the ex-Gaussian distribution within the range of parameter values appropriate for modeling RT. As is illustrated in Figure 2, ex-Wald and ex-Gaussian distributions with equal parameters have very similar values, and in fact, they have virtually identical values at all percentiles above approximately 5%–10%. Given that the ex-Gaussian provides a satisfactory approximation to RT distribution (see, e.g., Hockley, 1984; Hohle & Gholson, 1968; Luce, 1986), the ex-Wald must do so too. Moreover, the fact that the lower tail of the ex-Gaussian extends below zero in principle suggests that, if anything, the ex-Wald is the better model for the shortest RTs.

Because of its clear theoretical advantages over the ex-Gaussian, I chose to model RTs with the ex-Wald distribution for the simulations in this article. Unlike Schwarz (2001), I will characterize each ex-Wald distribution in terms of three parameters analogous to those used with the ex-Gaussian—namely, the mean and standard deviation of the Wald component, μ_w and σ_w , and the mean of its exponential component, μ_e . For example, the ex-Wald with $\mu_w = 450$, $\sigma_w = 45$, $\mu_e = 100$ is the convolution of a Wald distribution having a mean and standard deviation of 450 and 45 msec, respectively, and an exponential distribution having a mean of 100 msec.

Simulation Conditions

As will be described in detail in this section, there were 216 sets of simulations defined by a factorial combination of four simulation parameters: 2 control condition distributions \times 3 effect loci \times 6 effect probabilities (i.e., values of P) \times 6 numbers of trials per condition. The first three of these factors determined the exact probability distributions of the control and experimental conditions being compared, yielding the factorial structure depicted in Figure 3. The fourth factor—number of trials per condition—simply reflected the amount of data collected.

Control condition distributions. Two different control condition distributions, both with overall mean RTs of 500 msec, were chosen to represent control conditions

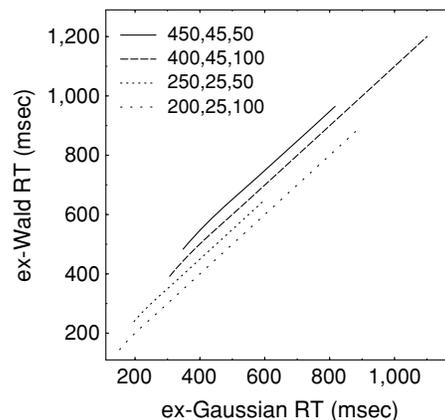


Figure 2. Quantile–quantile plots displaying the similarity of the ex-Wald and ex-Gaussian distributions. Each line displays the reaction time (RT) at a given percentile within the ex-Wald distribution plotted against the RT at the same percentile within the ex-Gaussian distribution. Within each line, the Wald and Gaussian components of the two distributions have equal means and standard deviations, and the exponential components have equal rates. The first two parameters associated with each line indicate the means and standard deviations of the Gaussian and Wald components, which were, in each case, identical to one another; the third parameter is the mean of the exponential component. If the ex-Gaussian and ex-Wald distributions were identical, these lines would be perfectly straight and have slopes of one, and the deviations from this ideal are not large. Thus, the ex-Wald distribution is numerically quite similar to the ex-Gaussian with the corresponding parameter values. The upper three lines have been shifted upward in successive 50-msec increments to improve their visibility.

with relatively lower versus higher degrees of RT skew. The PDFs of these two distributions are shown as the thick lines in the left (lower skew) and right (higher skew) columns of Figure 3. Specifically, these distributions were the ex-Wald with $\mu_w = 450$, $\sigma_w = 45$, and $\mu_e = 50$, and the ex-Wald with $\mu_w = 400$, $\sigma_w = 40$, and $\mu_e = 100$, respectively.

Effect loci. In all the simulations, the true mean RT in the experimental condition was 550 msec, representing a 50-msec effect on mean RT, relative to the control condition. Three different effect loci were studied. The effect could result entirely from an increase in the mean of the Wald component of the ex-Wald, yielding a *Wald locus* of the effect, as is shown in panels A and B of Figure 3. Alternatively, it could result entirely from an increase in the mean of the exponential component of the ex-Wald (i.e., have an *exponential locus*), as is shown in panels E and F. Finally, it could result from equal increases in the means of both components of the ex-Wald (i.e., *split locus*), as is shown in panels C and D.³

Effect probabilities. For each different control condition distribution, experimental conditions were formed with six different probabilities of effect-present experimental trials—namely, $P = .05, .10, .25, .50, .75$, or 1.00.

The simulations with $P = 1.00$ represent uniform effects, and the PDFs for these experimental conditions

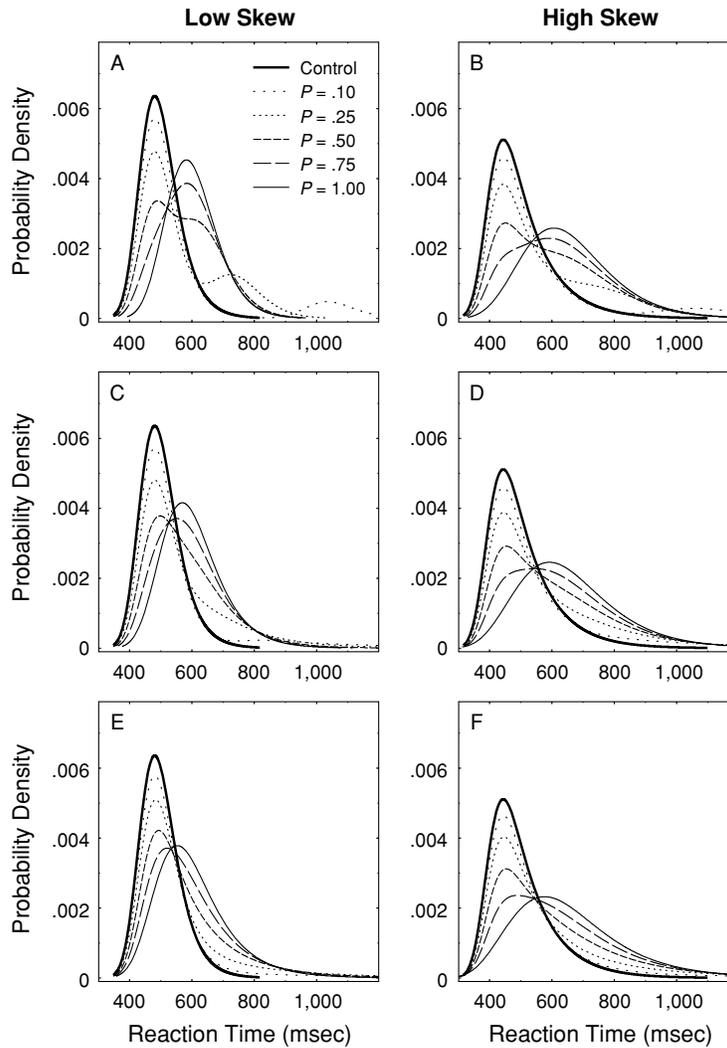


Figure 3. The probability density functions (PDFs) of reaction time (RT) underlying the simulations examining the likelihood ratio test. Each simulation compared the control condition, shown as a thick line, against one of the experimental conditions (thin lines) within the same panel. The control condition ex-Wald PDF in the left panels was chosen to have lower skew, and the one in the right panels was chosen to have higher skew. These control conditions, which are identical for all three panels on the left and for all three panels on the right, are repeated across panels for ease of comparison with the experimental distributions. The experimental effect increased the mean of the Wald component of the ex-Wald in panels A and B, it increased the mean of the exponential component in panels E and F, and it increased the means of both components equally in panels C and D. Different simulations used different values of the probability that the effect is present in an experimental trial, P , as indicated by the curve parameter. With $P = 1.00$, the PDFs in the experimental conditions were ex-Wald distributions. With $P < 1.00$, the PDFs in the experimental distribution were mixtures of two ex-Wald distributions, one of which was identical to that of the control condition and the other of which was adjusted to produce an overall effect size of 50 msec on the mean RTs.

are depicted as the thin solid lines in Figure 3. Consider panel A, for example. In some simulations, control condition RTs were sampled from the ex-Wald with $\mu_w = 450$, $\sigma_w = 45$, and $\mu_e = 50$ (PDF shown in thick solid line), and experimental condition RTs were sampled from the ex-

Wald with $\mu_w = 500$, $\sigma_w = 45$, and $\mu_e = 50$ (PDF shown in thin solid line). As will be described further in the section on Simulation Procedure, the required numbers of RTs (depending on number of trials per condition, discussed next) were randomly sampled from each of these

two distributions, uniform and mixture model parameters were estimated, and λ was computed. Ideally, the likelihood ratio test should not lead the experimenter to reject the null hypothesis in simulations with uniform effects, so the Type I error rate of the procedure can be estimated from the simulations with $P = 1.00$.

The other values of P (i.e., $P < 1.00$) represent mixture effects with different effect-present probabilities. With an effect probability of $P = .10$, for example, the experimental effect was present (i.e., RT was lengthened) in 10% of all the experimental trials, and it was absent (i.e., RT came from the same distribution as in the control condition) with a probability of $1 - P = .90$. Thus, for $P < 1.00$, trials in the experimental condition were a mixture of some trials equivalent to those from the control condition and some trials from an effect-present distribution, the exact construction of which will be described next.

Effect-Present Distributions

Given the well-known properties of mixture distributions (see, e.g., Everitt & Hand, 1981; McLachlan & Basford, 1988; Titterton, Smith, & Makov, 1985; Townsend & Ashby, 1983; Yantis, Meyer, & Smith, 1991), the distribution of RTs for effect-present trials was uniquely determined by the combination of the control condition distribution, the effect probability, and the effect locus. For example, note that the true mean RT of the effect-present experimental trials must be inversely proportional to P in order to maintain the constant overall mean of 550 msec in the experimental condition. This stems from the fact that the true mean in the experimental condition, μ_{e*} , is a weighted average of the true mean in the control condition, μ_c , and the true mean RT for effect-present trials in the experimental condition, μ_{e*} :

$$\mu_e = (1 - P) \cdot \mu_c + P \cdot \mu_{e*}. \quad (7)$$

With effect probabilities of $P = .05, .10, .25, .50, .75$, and 1.00 , then, the corresponding effect-present mean RTs were required to be 1,500, 1,000, 700, 600, 567, and 550 msec, respectively, to produce the desired overall mean of 550 msec in each experimental condition.

Effect-present experimental RTs were also assumed to come from ex-Wald distributions, and the parameters of this distribution were adjusted to produce the required mean for each value of P . Depending on the effect locus (i.e., Wald, exponential, or split), the parameters of the Wald component, the exponential component, or both were adjusted to give the required means, with the standard deviation of the Wald component remaining identical to that in the control condition.⁴ The dotted and dashed lines in Figure 3 display the PDFs of the resulting true mixture distributions used to simulate the experimental conditions, and the corresponding CDFs are shown in Figure 4.

Numbers of trials. For each of the different combinations of control and experimental conditions described, simulations were run of experiments with six different numbers of trials per condition: $N = 25, 50, 100, 250, 500$, or $1,000$. The number of trials per condition in an

actual experiment would be selected by the experimenter, of course, as a function of various practical constraints limiting data collection.

Simulation Procedure

An identical procedure was used to simulate 3,000 experiments under each of the 216 combinations of simulation parameters defined by the factorial combination of 2 control condition distributions \times 3 effect loci \times 6 effect probabilities (i.e., values of P) \times 6 numbers of trials per condition. First, N RTs were randomly generated within each condition (i.e., control and experimental), with RTs from each condition sampled from the distribution appropriate for the given simulation parameters, as has been described above. Second, the simplex algorithm (Rosenbrock, 1960) was used to find maximum likelihood estimates of the parameters for the uniform effect model, and the associated likelihood value according to this model, Γ_u . Third, the simplex algorithm was also used to find maximum likelihood estimates of the parameters for the mixture effect model, and the associated likelihood value according to this model, Γ_m .⁵ Fourth, the value of the likelihood ratio test, λ , was computed from these two likelihood values, using Equation 6.

Simulation Results

Figure 5 summarizes the results with respect to the outcomes of the likelihood ratio tests, plotting the probability of rejecting the null hypothesis of a uniform effect at the .05 significance level (i.e., the probability of concluding that there is a mixture effect) as a function of the simulation parameters. Note that when the true effect probability is $P = 1.0$, the plotted probability indicates the Type I error rate of the test. In contrast, when the true effect probability is $P < 1.0$, the plotted probability indicates the test's power.

Inspection of Figure 5 reveals several important properties of the proposed likelihood ratio test. First, the power of the test increased rather systematically with increases in the number of observations per condition, as would be expected. Second, examination of the conditions with $P = 1.0$ indicates that the Type I error rate was appropriately low, being at most slightly above the nominal 5% level and often below that. Thus, although the test statistic has theoretically only an approximate χ^2 distribution, it appears that the test leads to an incorrect rejection of the null hypothesis in approximately the correct proportion of samples.

Third, the power of the test increased almost monotonically as the true effect probability decreased. This pattern is intuitively reasonable, because the effect-present mean RT is more dissimilar from the control mean RT when the effect probability is small, as was discussed in connection with Equation 7. Intuitively, increased separation of these two means makes it easier to distinguish the two types of trials within the mixture. At the smallest effect probabilities, power may drop off because of the increased chance of an experimental sample containing no effect-present RTs.

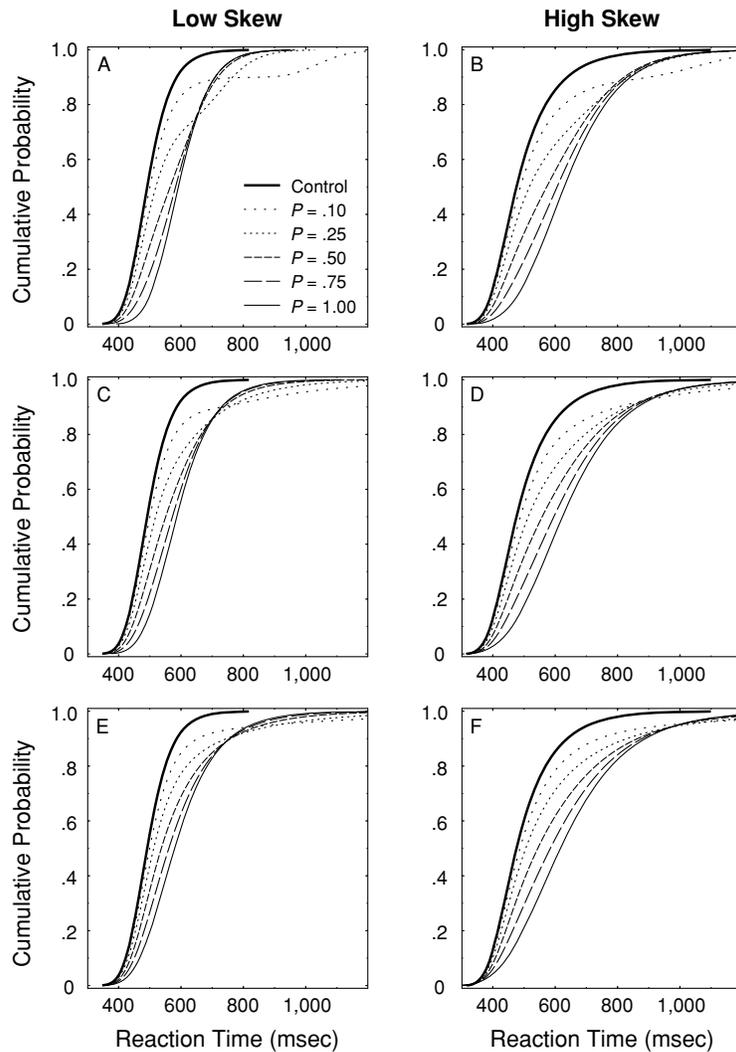


Figure 4. The cumulative probability density functions (CDFs) of reaction time (RT) underlying the simulations examining the likelihood ratio test. The panels and parameters are identical to those in Figure 3.

Fourth, the locus of the effect had quite a large influence on the power of the test. Power was clearly largest when the experimental manipulation influenced the Wald component of the ex-Wald distribution (panels A and B of Figure 5), intermediate when the manipulation influenced the exponential component (panels E and F), and smallest when the manipulation influenced both components equally (panels C and D). Fortunately, even in the worst case of the split locus, power was reasonably good for the most important range of $P < .2$, especially with at least 100 observations per condition. In contrast to the large influence of the effect locus, there was relatively little influence of the overall skewness of the RTs, as indicated by the similarity of the left and the right panels.

Figure 6 shows the estimated mixture probability, \hat{P} , averaged across iterations for each simulation condition. These indicate that the estimate is remarkably unbiased

when the experimental effect influences the Wald component of the ex-Wald distribution (panels A and B), except for a slight underestimation of the true mixture probabilities near one. When the effect influences both components equally (panels C and D), there is a small positive bias when $P < .8$, with \hat{P} tending to overestimate the true value. This overestimation is somewhat more pronounced when the effect influences the exponential component (panels E and F). Despite these small biases, it appears that the technique may be used to get a reasonably good estimate of the true mixture probability, especially if the number of observations per condition is fairly large.

Aggregation of Results Across Experimental Participants

So far, I have considered the likelihood ratio test for mixture effects using a single data set with observations

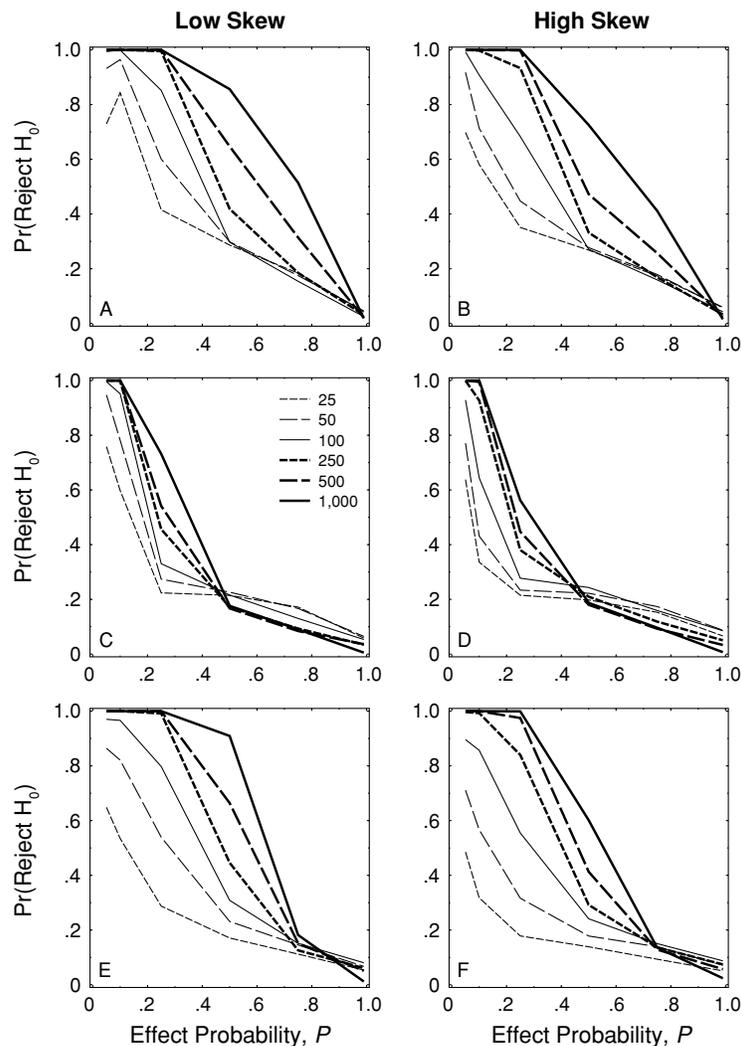


Figure 5. Probability of rejecting the null hypothesis that the true effect is a uniform effect (i.e., of concluding that the true effect is a mixture effect; vertical axis) as a function of the true effect probability (horizontal axis), the sample size (line parameter), and the different simulation conditions depicted in Figures 3 and 4.

from each of two conditions. Often, however, RT researchers test many experimental participants for a relatively short period each, using a within-subjects design in which each participant is tested in both of the conditions of interest. It is therefore useful to consider how the test for mixture effects could be aggregated across participants in such a within-subjects design.

It is well known in the statistical literature (e.g., Dudewicz & Mishra, 1988) that the sum of independent χ^2 random variables is itself χ^2 , with the degrees of freedom of the sum equal to the sum of the degrees of freedom values of the individual summands. In the present situation, then, the observed value of the likelihood ratio test can be computed for each of S participants and simply summed across participants. The distribution of the sum

would be expected to have an approximate χ^2 distribution with S degrees of freedom. Thus, an observed sum could be checked against the critical χ^2 value from this distribution to see whether the null hypothesis of a uniform effect could be rejected when aggregating across participants.

Figure 7 shows the results of simulations employing this aggregation procedure for the case of aggregating across 20 cases. In each iteration of this simulation, 20 of the 3,000 simulated cases for a given simulation condition were randomly chosen without replacement. The sum of these 20 observed λ values was computed, and this sum was checked for significance at the .05 significance level. More specifically, the null hypothesis of a uniform effect was rejected if this sum exceeded the critical value of 31.41 from the χ^2 distribution with 20 degrees of freedom.

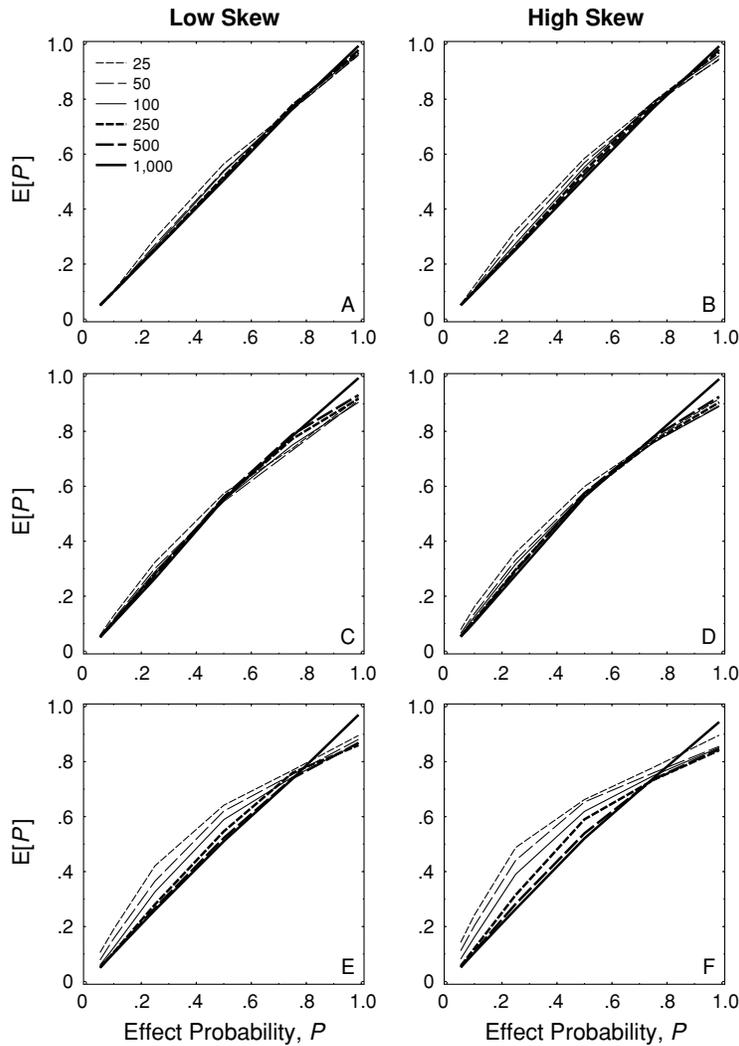


Figure 6. Mean estimated mixture probability, \hat{P} (vertical axis), as a function of the true effect probability (horizontal axis), the sample size (line parameter), and the different simulation conditions depicted in Figures 3 and 4.

This aggregation procedure was iterated 30,000 times for each simulation condition, and the proportions of significant sums are shown in the figure.

The results in Figure 7 indicate that the aggregated version of the likelihood ratio test has good statistical properties. First, as was expected, it is clearly possible to increase the statistical power of the test by aggregating across participants. That is, for any given set of simulation conditions (i.e., value of P , number of trials per condition, etc.), the probability of correctly rejecting the null hypothesis of a uniform effect is substantially larger when aggregating across 20 participants than it was for each individual participant, as can be seen by comparing Figures 5 and 7. Second, the Type I error rate is inflated little, if any, by the aggregation procedure, with approximately 5% rejections of a true null hypothesis in the case of $P = 1$ under all conditions.

DISCUSSION

The proposed likelihood ratio test is a promising new tool for gaining additional information from RT distributions beyond that available from mean RT. Specifically, the test can be used to investigate whether an experimental manipulation has a uniform effect, which influences performance on all trials, or a mixture effect, which influences performance on only a subset of trials. For this test, the null hypothesis is instantiated in a model with a uniform effect. If the fit of the alternative mixture model is sufficiently superior to that of the uniform model, the null hypothesis is rejected, and it is concluded that the manipulation has a mixture effect.

Under conditions similar to those of many RT experiments, the test has both an appropriate Type I error rate and good power to detect mixtures when they are present,

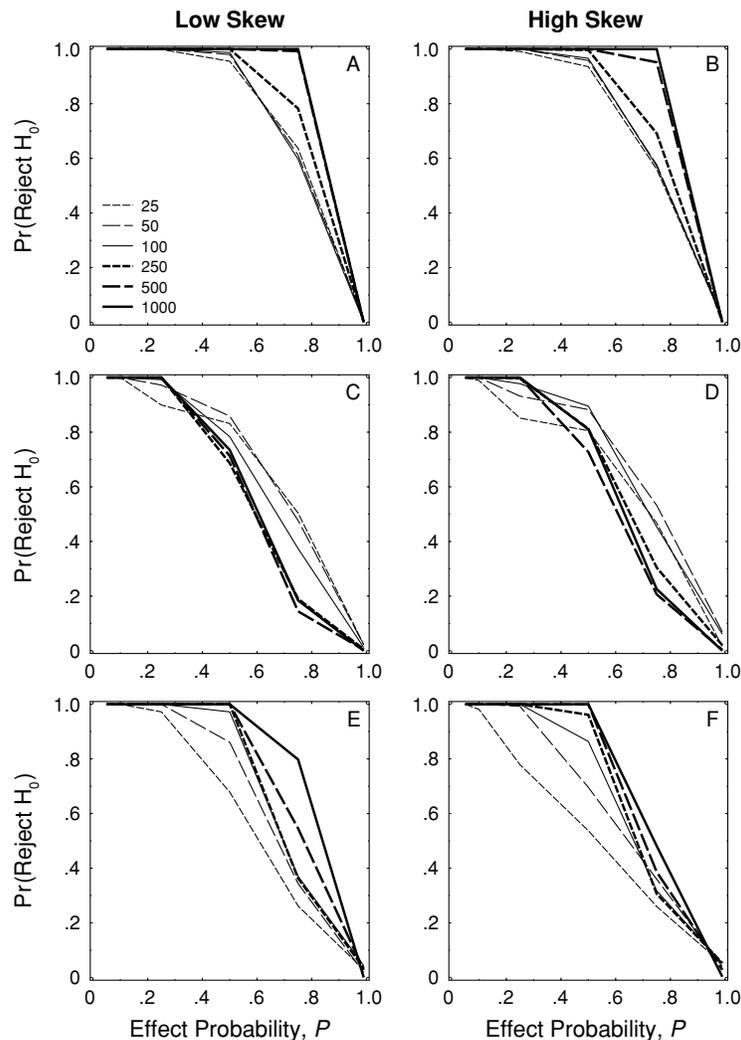


Figure 7. Probability of rejecting the null hypothesis based on an aggregation of λ values across 20 independent experimental participants. This figure is exactly parallel to that in Figure 5, except that the null hypothesis was tested via the aggregated test.

especially when there are many trials or results are aggregated across a number of experimental participants. These results provide encouragement that the test would lead to appropriate decisions about effect type with real data sets. It is especially encouraging that power is highest when the effect probability is relatively low, because these are the cases of most theoretical interest in RT research. That is, the interpretation of an experimental effect will usually depend greatly on whether it is present in a small minority of trials, as opposed to all trials, but it will depend relatively little on whether the effect is present in most trials versus all trials. In addition, the maximum likelihood estimates of the mixture probability generally show fairly small biases, suggesting that it may be possible to estimate this probability reasonably accurately when a mixture effect is present.

Comparison With Other Methods

To my knowledge, RT researchers interested in distinguishing between uniform and mixture effects have previously made only informal comparisons of RT distributions. One type of informal comparison is to check for a bimodal frequency distribution in the experimental condition. This approach can be illustrated by considering the dotted and dashed lines in Figure 3, which display the PDFs of the resulting true mixture distributions used to simulate the experimental conditions. For the low-skew control condition and a Wald locus of the effect, as shown in panel A, the mixture character of these distributions is clearly suggested by the bimodality of the PDF when $P < .5$. Unfortunately, bimodality is not a reliable diagnostic of mixture effects, because mixtures of unimodal PDFs are rarely bimodal (Townsend & Ashby, 1983, p. 269). For

the examples shown in Figure 3, bimodality is clearly evident only in panel A; with higher skew or when the effect does not have a Wald locus, the mixture distributions are sufficiently smeared to conceal the underlying mixture. This suggests that observed PDFs will not generally be sufficient to determine whether an experimental effect is a uniform or a mixture effect. Thus, although bimodality is clear evidence for a mixture effect, the absence of bimodality is not strongly indicative of a uniform one.

Another informal method is to compare the CDFs of the RTs observed in the experimental and control conditions. As is illustrated in panel D of Figure 1, the intuitive expectation is that a uniform effect should basically produce a shift of the CDF in the experimental condition, yielding approximately parallel CDFs for the two conditions. In contrast, a mixture effect should produce CDFs that fan out from a common minimum, as is illustrated in panel E. In practice, this difference is most easily checked by seeing whether the experimental effect is larger for high percentiles than for low ones.

Close inspection of the CDFs shown in Figure 4 suggests, however, that there may be problems with the practice of checking for mixtures by seeing whether the CDFs diverge at the high end. Although the mixture character is evident in some cases as a tendency for the CDF to have a visible flattening or plateau at some point below the maximum, such flattening is not generally sufficient to detect mixtures, because it is not present in all cases. Worse yet, there is visible divergence to some extent for almost all pairs of experimental condition and control condition CDFs, even for most of the uniform experimental conditions. Thus, checking for divergence might tend to suggest the presence of a mixture effect even when the true effect was uniform.

An Actual Experimental Example

To illustrate the method in the analysis of data from an actual experiment, I will present a reanalysis of results reported by Franz and Miller (2002). In brief, they compared the performance of university students, Parkinson's patients, and age-matched controls in a go versus no-go task with visual stimuli. In each trial, the participants were given a precue to indicate the probability that the go response would be required in that trial. Two different precues indicated go probabilities of either .8 or .2. For all three groups of participants, responses were approximately 35 msec faster following the cue indicating high go probability than following the cue indicating low go probability.

Franz and Miller (2002) interpreted the effect of cued go probability as suggesting greater response readiness when go probability was high than when it was low, as had others before them (e.g., Mattes, Ulrich, & Miller, 1997). It is possible to refine this interpretation, however, by considering the distinction between uniform and mixture effects. Specifically, it is possible that the participants consistently reached only a relatively low level of response preparation in the low-probability condition and, thus, produced consistently slowed responses (i.e., a uniform effect). Alternately,

it is possible that participants sometimes prepared fully in the low-probability condition—just as they did in the high-probability condition—but did not prepare in other trials (i.e., a mixture effect). For example, the participants might have failed to prepare in half of the trials with a low-probability cue. This would have produced a 35-msec effect on mean RT if failing to prepare slowed responses by 70 msec.

To distinguish between these two possibilities, I analyzed the RTs obtained from each participant, using the likelihood ratio test presented here. I included only those participants for whom the mean RT in the high-probability condition was numerically smaller than the mean RT in the low-probability condition, reasoning that the participants not showing the effect could provide no information as to its nature.

For the age-matched control group, the results provided little evidence of a mixture effect. None of the 8 participants included in the analysis yielded a significant λ value, and the aggregated test did not approach significance ($p > .25$). Across participants, the average estimated \hat{P} value was .75.

For the Parkinson's patients and university students, however, there were clear signs of mixture effects. The λ value was significant ($p < .05$) for 7 of the 17 students included in the analysis and for 3 of the 10 Parkinson's patients. The aggregated test statistic was highly significant ($p < .01$) for both groups, and the average estimated \hat{P} value was approximately .60 for the students and .40 for the patients. Thus, for these participants, the mixture model fitted enough better than the uniform model to suggest that go probability did not have an effect in every trial.

It remains for further research on response preparation to explore the implications of this apparent mixture effect in university students and in Parkinson's patients, as well as its contrast with apparently uniform effects in age-matched controls. Nonetheless, this example illustrates how the proposed new mixture test, like other RT distributional analyses, can provide information beyond that available in mean RT.

Limitations

An important limitation of the proposed test is that it can be used only if the researcher is able to specify the form (e.g., ex-Wald, ex-Gaussian, log normal, gamma, etc.) of the underlying RT distributions. Fortunately, with sufficient observations per condition, the researcher can use standard techniques to evaluate the fit of alternative theoretical models to the observed RT distributions, thus letting the data guide the choice of underlying distribution. Even in situations in which the researcher has too few observations per condition to obtain such guidance, however, this limitation does not seem very severe. One reason is that there is considerable prior research to indicate that RT distributions commonly have a form that is well approximated by the ex-Gaussian and, by extension, by the rather similar ex-Wald. A second reason is that informal simulations not reported here indicate that the test is robust with

respect to the underlying distributional assumption. That is, the test still works quite well even if the wrong underlying distribution is specified (e.g., there are true log normal distributions of RTs, but the researcher assumes that the distributions are ex-Walds). To check on robustness, I carried out simulations analogous to those whose results are presented in Figures 5–7, using RTs generated from ex-Gaussian distributions, gamma distributions, and log normal distributions. In all cases, the test was computed using the incorrect assumption that the underlying distributions were ex-Walds. Despite the erroneous assumption, the test always had high power when a mixture effect was present and a low Type I error rate when a uniform effect was present, although in these cases the estimates of P had larger biases than are shown in Figure 6.

Another inherent limitation of the test is that an experimental manipulation with a uniform effect could, *in principle*, produce an observable RT distribution with a shape that just happened to match that of the alternative mixture model. In that case, the researcher would most likely incorrectly conclude that a mixture effect was present, committing a Type I error. This limitation also does not seem too serious, however, because most observed RT distributions are well fit by known approximation distributions (e.g., ex-Gaussian) that do not mimic mixture distributions. Moreover, even if the true RT distribution had some unusual form in an experimental condition with a uniform effect, it would be quite a coincidence if that distribution just happened to have a form consistent with a mixture distribution involving some proportion of observations from the control condition.

One limitation of the aggregated test should also be mentioned. This test treats the different participants in the experiment as the entire population of interest, with trials representing the random component, rather than treating the participants as a random sample from the population of interest. Formally, then, a significant result of an aggregated test should be regarded as support for the conclusion that there is a mixture effect *for at least 1 participant within the current group*, not necessarily for every participant in the whole population. A fairly large sample with relatively many trials per participant might be needed to estimate precisely the number of participants in the population for whom there is a mixture effect.

Possible Extensions

The likelihood ratio test discussed here could be extended in at least two ways, although it is beyond the scope of the present article to carry out simulations examining these extensions. One extension involves interval estimation of P , and the other involves consideration of other types of experimental effects.

Interval estimation of P . Although I have emphasized the comparison of uniform and mixture effects via hypothesis testing, another data-analytic approach would be to look for an interval estimate—analogue to a confidence interval or fiducial interval (Wang, 2000)—of the true probability that an effect is present in an experimental trial (i.e., P). The null hypothesis of a uniform effect

would be rejected if and only if this interval excludes the point $P = 1$; in that case, of course, the interval estimate contains more specific information about the true value of P .

To compute an interval estimate of P from a single set of control and experimental observations, the researcher could begin by fitting the unconstrained mixture model used in the likelihood ratio test. This is, by definition, the mixture model yielding the highest likelihood of the observed data, with an associated point estimate \hat{P} of P . The researcher can then fit constrained mixture models in which P is estimated by $\hat{P}_u > \hat{P}$, for various values of \hat{P}_u . For some value of \hat{P}_u , the fit of constrained model will be significantly worse than the fit of the unconstrained model, according to a likelihood ratio test. The lowest value \hat{P}_u at which this occurs could be taken as the upper end of the interval estimate for P . The lower end of the interval estimate could be found analogously using values of $\hat{P}_l < \hat{P}$. It is beyond the scope of this article to investigate the statistical properties of this interval estimation procedure, although it does seem to have reasonable face validity.

Alternatively, one might want to compute a confidence interval for the population average value of P by aggregating observations across many experimental participants. At present, the best way to do this would seem to be to compute a confidence interval for the mean of the \hat{P}_i values, where \hat{P}_i is the estimated effect probability for participant i . This method essentially treats the individual-participant \hat{P}_i values like any other summary measure of individual performance, using between-participant central tendency and variability to estimate the true population value.

Additive effects. The likelihood ratio test can also be extended to handle certain other specific types of experimental effects, as long as the distribution of effect-present RTs can be specified exactly. For RT research, the most plausible extension would be to consider additive experimental effects in which each experimental RT is conceived of as a control RT plus a random increment having PDF h . Such an additive conception of the experimental effect would seem appropriate, for example, when comparing two tasks requiring different numbers of mental operations (e.g., simple RT vs. choice RT, as considered by Donders, 1868/1969; visual search and memory search through different numbers of items, as considered by Neisser, 1963, and Sternberg, 1966).

In cases of additive effects, the distribution of effect-present RTs in the experimental condition can be represented as the convolution of the control distribution with the increment distribution, so that

$$f_{\text{eu}}(t) = \int_0^t h(x)f_{\text{cu}}(t-x)dx$$

and

$$f_{\text{e}}(t) = \begin{cases} \int_0^t h(x)f_{\text{cm}}(t-x)dx & \text{with probability } P \\ f_{\text{cm}}(t) & \text{with probability } 1 - P. \end{cases}$$

Using these PDFs, maximum likelihood parameter estimates can be obtained, and λ can be computed as before. Further simulations would be needed, however, to examine the statistical properties of λ in such cases.

SOFTWARE

The likelihood ratio test described here can be computed with the program MIXTEST, which can be freely downloaded via links at the author's Web page psy.otago.ac.nz/miller/index.htm. This program can be used both for the analysis of data and for the simulation of the likelihood ratio test under a wide variety of conditions. In particular, the program can be used with many different distributions based on those of CUPID (Miller, 1998). These distributions include the ex-Wald, ex-Gaussian, gamma, and log normal, which might be useful in RT research, as well as the normal, exponential, uniform, and many others, some of which would be most appropriate outside the context of RT research. MIXTEST runs in a command line window under the Windows operating system (Version 98 or later), and it reads data and program control information from plain-text files. Full documentation is supplied with the program.

REFERENCES

ALLEN, P. A., SMITH, A. F., VIRES-COLLINS, H., & SPERRY, S. (1998). The psychological refractory period: Evidence for age differences in attentional time-sharing. *Psychology & Aging*, **13**, 218-229.

BORGER, R. (1963). The refractory period and serial choice-reactions. *Quarterly Journal of Experimental Psychology*, **15**, 1-12.

CHEN, H., & CHEN, J. (2001). The likelihood ratio test for homogeneity in the finite mixture models. *Canadian Journal of Statistics*, **29**, 201-216.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics*, **25**, 573-578.

CHHIKARA, R. S., & FOLKS, J. L. (1989). *The inverse Gaussian distribution: Theory, methodology, and applications*. New York: Dekker.

COX, D. R., & MILLER, H. D. (1965). *The theory of stochastic processes*. London: Chapman & Hall.

DAWSON, M. R. W. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments, & Computers*, **20**, 54-57.

DIEDERICH, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, **39**, 197-215.

DIXON, P. (2003). The *p*-value fallacy and how to avoid it. *Canadian Journal of Experimental Psychology*, **57**, 189-202.

DONDERS, F. C. (1969). Over de snelheid van psychische processen [On the speed of mental processes] (W. G. Koster, Trans.). In W. G. Koster (Ed.), *Attention and performance II* (pp. 412-431). Amsterdam: North-Holland. (Original work published 1868)

DUDEWICZ, E. J., & MISHRA, S. N. (1988). *Modern mathematical statistics*. New York: Wiley.

ERIKSEN, B. A., & ERIKSEN, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16**, 143-149.

EVERITT, B. S., & HAND, B. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.

FAZIO, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (Review of personality and social psychology, Vol. 11, pp. 74-97). Newbury Park, CA: Sage.

FRANZ, E. A., & MILLER, J. O. (2002). Effects of response readiness on reaction time and force output in people with Parkinson's disease. *Brain*, **125**, 1733-1750.

GOTTLLOB, L. R. (2004). Location cuing and response time distributions in visual attention. *Perception & Psychophysics*, **66**, 1293-1302.

GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

HEATHCOTE, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments, & Computers*, **28**, 427-445.

HEATHCOTE, A. (2004). Fitting Wald and ex-Wald distributions to response time data: An example using functions for the S-PLUS package. *Behavior Research Methods, Instruments, & Computers*, **36**, 678-694.

HEATHCOTE, A., POPIEL, S. J., & MEWHORT, D. J. K. (1991). Analysis of response-time distributions: An example using the Stroop task. *Psychological Bulletin*, **109**, 340-347.

HOCKLEY, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 598-615.

HOEL, P. G. (1962). *Introduction to mathematical statistics* (3rd ed.). New York: Wiley.

HOHLE, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, **69**, 382-386.

HOHLE, R. H., & GHOLSON, B. (1968). Choice reaction times with equally and unequally probable alternatives. *Journal of Experimental Psychology*, **78**, 95-98.

JUHEL, J. (1993). Should we take the shape of reaction time distributions into account when studying the relationship between RT and psychometric intelligence? *Personality & Individual Differences*, **15**, 357-360.

LETH-STEENSEN, C., ELBAZ, Z. K., & DOUGLAS, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychologica*, **104**, 167-190.

LUCE, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford: Oxford University Press.

MATTES, S., ULRICH, R., & MILLER, J. O. (1997). Effects of response probability on response force in simple RT. *Quarterly Journal of Experimental Psychology*, **50A**, 405-420.

MCLACHLAN, G. J., & BASFORD, K. E. (1988). *Mixture models*. New York: Dekker.

MEWHORT, D. J. K., BRAUN, J. G., & HEATHCOTE, A. (1992). Response time distributions and the Stroop task: A test of the Cohen, Dunbar, and McClelland (1990) model. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 872-882.

MEYER, D. E., OSMAN, A. M., IRWIN, D. E., & YANTIS, S. (1988). Modern mental chronometry. *Biological Psychology*, **26**, 3-67.

MEYER, D. E., YANTIS, S., OSMAN, A. M., & SMITH, J. E. K. (1985). Temporal properties of human information processing: Tests of discrete versus continuous models. *Cognitive Psychology*, **17**, 445-518.

MILLER, J. O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, **14**, 247-279.

MILLER, J. [O.] (1998). Cupid: A program for computations with probability distributions. *Behavior Research Methods, Instruments, & Computers*, **30**, 544-545.

NEISSER, U. (1963). Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology*, **76**, 376-385.

PASHLER, H. E. (1991). Shifting visual attention and selecting motor responses: Distinct attentional mechanisms. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 1023-1040.

PASHLER, H. E. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, **116**, 220-244.

PASHLER, H. E., & JOHNSTON, J. C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *Quarterly Journal of Experimental Psychology*, **41A**, 19-45.

RAO, C. R. (1965). *Linear statistical inference and its applications*. New York: Wiley.

RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.

- RATCLIFF, R., GOMEZ, P., & MCKOON, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, **111**, 159-182.
- RATCLIFF, R., & ROUDER, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347-356.
- ROSENBRACK, H. H. (1960). An automatic method for finding the greatest or least value of a function. *Computer Journal*, **3**, 175-184.
- ROUDER, J. N., & SPECKMAN, P. L. (2004). An evaluation of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review*, **11**, 419-427.
- SCHWARZ, W. (1993). A diffusion model of early visual search: Theoretical analysis and experimental results. *Psychological Research*, **55**, 200-207.
- SCHWARZ, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology*, **38**, 504-520.
- SCHWARZ, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers*, **33**, 457-469.
- SCHWARZ, W. (2002). On the convolution of inverse Gaussian and exponential random variables. *Communications in Statistics: Theory & Methods*, **31**, 2113-2121.
- SMITH, M. C. (1969). The effect of varying information on the psychological refractory period. *Acta Psychologica*, **30**, 220-231.
- SMITH, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, **44**, 408-463.
- STERNBERG, S. (1966). High-speed scanning in human memory. *Science*, **153**, 652-654.
- TITTERINGTON, D. M., SMITH, A. F. M., & MAKOV, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
- TOMBU, M., & JOLICŒUR, P. (2002). All-or-none bottleneck versus capacity sharing accounts of the psychological refractory period phenomenon. *Psychological Research*, **66**, 274-286.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- USHER, M., OLAMI, Z., & MCCLELLAND, J. L. (2002). Hick's law in a stochastic race model with speed-accuracy tradeoff. *Journal of Mathematical Psychology*, **46**, 704-715.
- VAN ZANDT, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, **7**, 424-465.
- WANG, Y. H. (2000). Fiducial intervals: What are they? *American Statistician*, **54**, 105-111.
- WELFORD, A. T. (1959). Evidence of a single-channel decision mechanism limiting performance in a serial reaction task. *Quarterly Journal of Experimental Psychology*, **11**, 193-210.
- WENGER, M. J., & TOWNSEND, J. T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *Journal of General Psychology*, **127**, 67-99.
- WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, **9**, 60-62.
- YANTIS, S., MEYER, D. E., & SMITH, J. E. K. (1991). Analyses of multinomial mixture distributions: New tests for stochastic models of cognition and action. *Psychological Bulletin*, **110**, 350-374.

NOTES

1. The requirement that the two competing models assume the same distributional forms for each of the two conditions is needed to ensure that the uniform model is a special case of the mixture model, as is a requirement of the χ^2 likelihood ratio test. Note that for the special case of $P = 1$, the mixture model likelihood given by Equation 5 reduces to

$$\Gamma_m = \prod_{i=1}^{N_c} f_{cm}(X_i) \times \prod_{i=1}^{N_c} f_{cm}(Y_i).$$

If $f_{cu} \equiv f_{cm}$ and $f_{eu} \equiv f_{cm}$, then this likelihood is equivalent to the likelihood under the uniform model (i.e., Equation 3). Thus, the identities of these two distributional pairs ensure that the two models are equivalent for the case of $P = 1$ and, hence, that the uniform model is a special case of the mixture model.

2. Exact specifications of the PDF and CDF of the ex-Wald distribution are available in Schwarz (2001, 2002) and will not be presented here.

3. A further refinement would be to differentiate the Wald-locus effects according to those influencing the drift rate of the diffusion process versus those influencing the cutoff criterion. For simplicity, I have not adopted this refinement in the present simulations. In analyzing actual RT data, though, a researcher would certainly want to compare parameter estimates across conditions to partition Wald-locus effects into these two possible types.

4. In a separate set of simulations, ex-Wald parameters were adjusted slightly differently, keeping a constant ratio of the standard deviation of the Wald component to the mean of this component (i.e., $\sigma_w/\mu_w = 0.1$). Analogous simulations were also carried out using ex-Gaussian distributions instead of ex-Walds. All of these simulations gave results almost identical to those presented here, so they will not be reported.

5. Like all hill-climbing algorithms, the simplex algorithm can converge on a local maximum likelihood, rather than on the global one required by Equation 6. This problem is usually not too serious in the analysis of real data, because researchers can start the algorithm at a variety of points and take the point of best convergence. To give the simplex algorithm the best chance of finding the global maximum in the present simulations, the parameters were always started at their true population values (i.e., at the true values used in generating the RTs). Furthermore, when the uniform model was fit to data generated with a mixture effect (i.e., $P < 1.0$), it was fit three times with starting values corresponding to each of the three possible effect loci.

(Manuscript received January 25, 2005;
accepted for publication March 28, 2005.)