

Assessing the evidence for response time mixture distributions

P. Dixon

Published online: 18 August 2012
© Psychonomic Society, Inc. 2012

Abstract I describe a technique for comparing two simple accounts of a distribution of response times: A mixture model and a generalized-shift model. In the mixture model, a target distribution is assumed to be a mixture of response times from two other (reference) distributions. In the generalized-shift model, the target distribution is assumed to be a quantile average of the reference distributions. In order to distinguish these two possibilities, quantiles for the target distribution are estimated from the quantiles of the reference distributions assuming either a shift or a mixture, and the predicted quantiles are used to calculate the multinomial likelihood of the obtained data. Monte Carlo simulations reported here demonstrate that the index is relatively unbiased, is effective with moderate sample sizes and modest spreads between the reference distributions, is relatively unaffected by changes in the number of bins or by data trimming, can be used with data aggregated across subjects, and is relatively insensitive to a range of subject variations in distribution shape and in mixture or shift proportion. As an illustration, the index is applied to the interpretation of three effects from distinct paradigms: residual switch costs in the task-switching paradigm, the psychological refractory period effect, and sequential effects in the Simon task. I conclude that the multinomial likelihood index provides a useful and easily applied tool for the interpretation of effects on response time distributions.

Keywords Response time · Distribution mixtures · Distribution shape

Electronic supplementary material The online version of this article (doi:10.3758/s13428-012-0244-7) contains supplementary material, which is available to authorized users.

P. Dixon (✉)
Department of Psychology, University of Alberta,
Edmonton, AB, Canada
e-mail: peter.dixon@ualberta.ca

It is uncontroversial that distributions of response times are mixtures. For example, across trials, subjects may have different response criteria, different strategies, different levels of motivation, and varying attentional states (e.g., Ollman, 1966). Similarly, performance varies as a function not only of the stimulus on the current trial, but also of the stimulus and response on the previous trial (e.g., Gratton, Coles, & Donchin, 1992) and of whether the previous response was correct or incorrect (e.g., Rabbitt, 1966). Usually, a description of the distribution of response times ignores many of these variations and simply pools all of the trials from these different possible states and strategies. However, it is easy to think of situations in which experimental manipulations could affect the nature of these mixtures. Temporal variables such as response–stimulus interval and stimulus onset asynchrony could easily affect the level of preparation; each type of stimulus could be processed in a distinct manner; the context of stimuli could lead subjects to use different strategies; and so on. Thus, under many circumstances, it is plausible to suppose that experimental manipulations have their effect on mean response time by affecting the nature of the response time mixture.

In contrast, most commonly used statistical techniques are designed to identify the mean of a distribution, on the basis of the assumption that the manipulations of interest shift the distribution from one location to another. For example, the general linear model as it is commonly developed assumes that the observations from different conditions all have a normal distribution with a common standard deviation, and thus differ only in their means (e.g., Winer, Brown, & Michels, 1991). Such statistical models have a straightforward interpretation in the theoretical analysis of response times: Manipulations that increase the mean response time do so because they increase the duration of one or more component mental processes. However, manipulations that affect distribution mixtures may also affect mean response time, and such an effect has a qualitatively

different theoretical interpretation. For example, it is commonly found that in simple reaction tasks, increasing the interval between a warning signal and the imperative stimulus decreases the mean response time (e.g., Klemmer, 1957). One might interpret this effect in terms of the duration of mental processing. In particular, it is plausible to suppose that increasing the warning foreperiod increases the level of preparation or arousal, and that this altered state leads to faster processing of the stimulus. Alternatively, one could interpret this effect in terms of a distribution mixture: Increasing the foreperiod increases the likelihood that subjects will respond with an anticipation or “fast guess” (Ollman & Billington, 1972). Thus, it can be theoretically critical to distinguish effects on location (i.e., a distribution shift) from effects on distribution mixture.

While distribution mixtures sometimes produce an obvious bimodality in the response time distribution, in many cases they do not. In particular, if the variability of the response time data is large relative to the magnitude of the experimental effects, it may be difficult to distinguish effects on shift from those on mixture simply from a casual inspection. Moreover, in many cases in which such a distinction is of crucial importance, there may be only modest amounts of data. Consequently, a tool for distinguishing these two possibilities that does not require large data sets could be of some value. The present article describes such a tool.

The plan of the article is as follows. First, I introduce some vocabulary and notation that can be used to describe mixture distributions. This vocabulary is used to introduce a class of experimental designs that can be used to identify possible mixtures. Second, I discuss a number of previous techniques for identifying mixtures. Mixture distributions have a number of interesting mathematical properties that make them particularly tractable, and these properties have been used with some frequency in a variety of contexts. The present technique builds on several of these. Third, I describe the technique itself, which I refer to as a “multinomial likelihood index.” The approach gains a great deal of power by comparing an effect on distribution mixture to a plausible alternative.

Terminology

A mixture distribution can be characterized as the probabilistic combination of two or more random variables. Here, I introduce some terminology for describing the simplest possible case of a distribution mixture in a response time experiment. In particular, I assume that one is interested in whether a given target condition can be described as a mixture of two other distributions. I refer to the potential sources of the mixture distribution as the “slow reference distribution” and the “fast reference distribution.” If the

target distribution is a mixture, the density function for the target distribution can be written as

$$f(t) = kr_f(t) + (1 - k)r_s(t),$$

where $r_f(t)$ is the density for the fast reference distribution, $r_s(t)$ is the density for the slow reference distribution, and k is the mixture probability. This characterization has a simple process interpretation: Assume that a subject has two ways to do a task, a fast process and a slow process. The target distribution, $f(t)$, results if the subject uses the fast process with probability k and the slow process with probability $1 - k$. The same simple relation also holds for the cumulative probability distributions:

$$F(t) = kR_f(t) + (1 - k)R_s(t).$$

I will assume in this development that these three distributions are ordered—in other words,

$$R_f(t) < F(t) < R_s(t).$$

This is not a difficult condition to achieve because one would normally be interested in considering a mixture distribution as a description of the target response times when the mean response time is intermediate between those for the reference conditions. The ordering relation would hold in such cases as long as the variances are not dramatically different. Moreover, when the fast and slow reference distributions are ordered, any mixture of the two will be in between. However, as discussed below, this is often not helpful in identifying mixtures, since simple changes in location can also produce that same ordering relation.

In applying this analysis, it is not necessary to assume that the reference distributions are themselves pure. Indeed, the same logic applies if the reference distributions differ only in the relative proportions of other processes. For example, suppose that a given task can be performed with either a fast process or a slow process and that the distributions of response times in these two cases are $r_f(t)$ and $r_s(t)$. Even if pure forms of these two distributions are unavailable, one may be able to identify a condition that is primarily made up of fast responses, yielding a mixture distribution such as

$$f_1(t) = k_1r_f(t) + (1 - k_1)r_s(t),$$

and a condition that is primarily made up of slow responses, yielding a mixture distribution such as

$$f_2(t) = k_2r_f(t) + (1 - k_2)r_s(t),$$

with $k_1 > k_2$. In that case, an intermediate mixture of those two reference distributions would yield

$$\begin{aligned} f_3(t) &= k_3r_f(t) + (1 - k_3)r_s(t) \\ &= kf_1(t) + (1 - k)f_2(t), \end{aligned}$$

where $k = (k_3 - k_2) / (k_1 - k_2)$. In other words, the intermediate mixture can be expressed as a mixture of two other mixtures, as long as the same reference distributions are involved for all three conditions (cf. Thomas, 1969). The net result is that in order to seek evidence for mixtures, one need not have conditions that embody pure processes; instead, one simply needs to be able to find conditions in which the relative proportions of the mixture components vary.

Previous work on mixtures

A variety of important properties of mixture distributions have been known for some time, and these properties have been exploited to identify mixture distributions. Clearly, if the reference distributions have comparable variances and are sufficiently different in the means, a mixture of the two distributions will be bimodal, and tests of bimodality can be used to identify possible mixtures (e.g., Silverman, 1981). More generally, the variance of a mixture will be larger than the weighted average of the reference distributions (e.g., Falmagne, Cohen, & Dwivedi, 1975):

$$\sigma_t^2 = k\sigma_f^2 + (1-k)\sigma_s^2 + k(1-k)(\mu_f - \mu_s)^2,$$

where μ_t , σ_t^2 , μ_f , σ_f^2 , μ_s , and σ_s^2 are the means and variances of the target, fast reference, and slow reference distributions, respectively. In other words, the variance of a mixture is a weighted average of the variances of the reference distributions plus an increment that depends on the difference in means of the reference distributions. If the reference distributions have similar variances and differ in location by a fair amount, this implies that the variance of a mixture will be substantially larger. Thus, a large variance is a qualitative indicator of a possible mixture. This result was used by Eriksen and Eriksen (1972) in a study of visual masking. More precisely, a mixture distribution will contain some observations from both the fast and slow reference distributions. Consequently, the minimum of a mixture will be similar to the minimum of the fast reference distribution, and the maximum will be similar to that of the slow reference distribution (Sternberg, 1973).

Luce (1986) provided a proof that the raw moments of a mixture distribution are a weighted average of the moments of the reference distributions, with the same weighting factor for all moments (see also Everitt & Hand, 1981). That is,

$$\int t^n f(t) dt = k \int t^n r_f(t) dt + (1-k) \int t^n r_s(t) dt. \quad (1)$$

Olds, Cowan, and Jolicœur (2000) used this property to assess the evidence for mixture distributions in the context of a visual search task with advance information. On the

basis of their results, they rejected the null hypothesis that the weighting factor was the same for the first to the fourth moments, and thus rejected a simple mixture model for their task.

A visually striking aspect of mixture distributions is the fixed-point property (Falmagne, 1968): Given two reference distributions with density functions that overlap and cross at some point, all mixtures of those reference distributions must also cross at that point. This observation has a simple proof. If the reference distributions cross at a point t_0 , then

$$r_f(t_0) = r_s(t_0) = r_0,$$

so

$$f(t_0) = kr_f(t_0) + (1-k)r_s(t_0) = kr_0 + (1-k)r_0 = r_0,$$

which is independent of k . This feature of mixtures has been used by a number of authors to assess evidence for mixtures. For example, Lupker and Theios (1975) used the approach to test the hypothesis that choice response time could be modeled by a two-state model in which subjects could be either prepared or unprepared for a particular stimulus. In their task, subjects made choice responses for either four or six possible stimuli and varied the probability of one particular “key” stimulus. Graphs of the observed distribution of response times for the key stimulus suggested that the fixed-point property held well for the four-stimulus case, but less accurately for the six-stimulus case. More recently, Brown, Lehmann, and Poboka (2006) used the fixed-point property to assess and reject the failure-to-engage account of residual switch costs in the task-switching paradigm. (These data and the Brown et al. conclusions are discussed in more detail later, in the section on applications.)

Luce (1986) described a more general property of mixtures that can be derived by eliminating the parameter k from equations based on Eq. 1:

$$F(t) = \alpha(t)\mu + \beta(t),$$

where

$$\alpha(t) = \frac{R_s(t) - R_f(t)}{\mu_s - \mu_f}$$

and

$$\beta(t) = R_f(t) - \alpha(t)\mu_f.$$

In other words, for any point t in the distributions, one can use the reference distributions to derive a linear relation that must hold for any possible mixture. Luce illustrated this relation with data collected by Falmagne (1965) and Lupker and Theios (1977). Thomas (1969) derived a nonparametric significance test of the null hypothesis that a distribution is a mixture, based on the distribution of possible orderings of triplets of observations from the reference and target distributions.

(This technique appears not to have been applied to psychological data.)

A test that is closely related to the present approach was proposed by Yantis, Meyer, and Smith (1991). In their technique, data from the reference and target distributions are partitioned into a set of data bins, and the numbers of observations from each condition that fall into these bins are calculated. These frequencies form a multinomial distribution. Moreover, if the target distribution is a mixture of the reference distributions, the predicted bin probabilities are related just as they are for densities. That is, for each bin i ,

$$p_i = kp_{fi} + (1 - k)p_{si},$$

where p_i is the multinomial probability for the bin in the target condition and p_{fi} and p_{si} are the multinomial probabilities for the bin in the fast and slow reference conditions, respectively. Given these relationships, it is possible to estimate the values of k and the p_{is} by maximizing the likelihood of the observed data. The maximum likelihood corresponding to these estimates can be used to calculate the goodness-of-fit statistic G^2 . Because G^2 has an asymptotic chi-square distribution, one can perform a significance test of the null hypothesis that the target distribution is a mixture. The hypothesis would be rejected if the statistic is larger than the usual significance testing criterion. (The Yantis et al. formulation is actually more general than described here and can be applied to situations in which there are several possible mixtures of any number of reference conditions.) As an illustration of the technique, Yantis et al. applied the approach to the priming data of Meyer, Yantis, Osman, and Smith (1985, Exp. 1). In that task, a word or nonword was presented as a prime for a left or right arrow stimulus. Yantis et al. tested the hypothesis that response times from trials with an intermediate prime–target stimulus onset asynchrony (SOA) were a mixture of a prepared state (represented by a long-SOA condition) and an unprepared state (represented by an unprimed condition). They failed to reject the null hypothesis of a mixture and concluded that the mixture account provided a viable interpretation of the task.

From my perspective, the previous approaches have a number of shortcomings. First, many of them require powerful experiments so that the form of the distribution can be estimated accurately. This is unfortunate because mixture distributions might provide a theoretically interesting interpretation of data from many studies with only moderate-sized designs. Second, many of the approaches reviewed here rely on a significance test, with a mixture distribution being the null hypothesis. By itself, failing to reject the null hypothesis, without further information about the circumstances of the test and the alternatives, provides only weak evidence in favor of that hypothesis (e.g., Cohen, 1990;

Sedlmeier & Gigerenzer, 1989). This suggests that a significance test may not be the best approach to describing the evidence that the target distribution is a mixture.

Moreover, in many situations the null hypothesis can be assumed to be incorrect in principle, at least in detail (e.g., Berkson, 1938; Cohen, 1990). This is likely to be the case in the application to mixture distributions, since it seems inevitable that the target distribution would deviate at least to some extent from a mixture of the reference distributions. Thus, the question addressed in a significance test is not whether the null hypothesis is false, but rather whether the experimental design is sufficiently powerful to detect the deviation. It would be more fruitful to ask, rather, whether it provides a better account than some well-specified alternative. That is the approach taken in the present developments.

Comparing mixtures and shifts

The best approach to assessing how useful a mixture model might be is to compare it to a competing alternative description of the distribution. Here, I describe a comparison technique, the multinomial likelihood index, to discriminate a mixture from what I term a “generalized shift.” I use the term “index” here to highlight the fact that the technique provides a measure of the evidence with respect to the two models and does not provide a test to reject a null hypothesis. In what follows, I first describe what I mean by a generalized-shift model and argue that this provides a compelling alternative to a mixture model. Then, I illustrate how both the mixture model and the generalized-shift model have a simple interpretation in terms of the cumulative distributions for the target and reference conditions. Finally, I discuss the mechanics of calculating the multinomial likelihood index.

Commonly, theoretical interpretations of response times presume that experimental factors affect the locations of response time distributions. In other words, factors are assumed to produce a distribution shift. Furthermore, most analyses of response time data make use of the general linear model (e.g., analysis of variance or multiple regression), and in using such techniques, one implicitly assumes that the effect of a manipulation is simply to shift the distribution of response times, without affecting its shape. These considerations suggest that the most obvious alternative to a mixture model is to assume that the target distribution is shifted some portion of the distance between the fast and slow reference distributions. This would seem to be the default assumption that is adopted when features of the distribution are not considered in any detail. However, it is well known that under many circumstances, experimental conditions that produce longer response times also produce

more variable response times (e.g., Green & Luce, 1971). Thus, a more accurate interpretation of the default assumption concerning effects on response time is that factors may affect both the location and scale of response times in a commensurate manner. The net result is that the target distribution should be somewhere in-between the fast and slow reference distributions in terms of both location and scale. I refer to this assumption as a “generalized-shift model.”

The competing predictions of a mixture model and a generalized-shift model are illustrated in Fig. 1. As stated before, a cumulative distribution function for a mixture can be described as a combination of the cumulative distribution functions for a fast and a slow reference distribution:

$$F(p) = kR_f(p) + (1 - k)R_s(p). \quad (2)$$

As can be seen in Fig. 1, this can be regarded as a vertical weighted average of the two reference distributions. In contrast, the generalized-shift model assumes that the target distribution is somewhere between the fast and slow reference distributions in both location and shape. This is readily captured by describing the target distribution as a quantile average of the fast and slow reference distributions:

$$F^{-1}(p) = kR_f^{-1}(p) + (1 - k)R_s^{-1}(p), \quad (3)$$

where F^{-1} and R^{-1} are quantile functions—that is, the inverse cumulative distribution functions that map cumulative probability to response time. It is easy to show that in this formulation, the mean of the target distribution is the weighted mean of the reference distributions. Jiang, Rouder, and Speckman (2004) demonstrated that when the reference distributions are part of the same location–scale families, the same is true of the variance. More generally, any property of

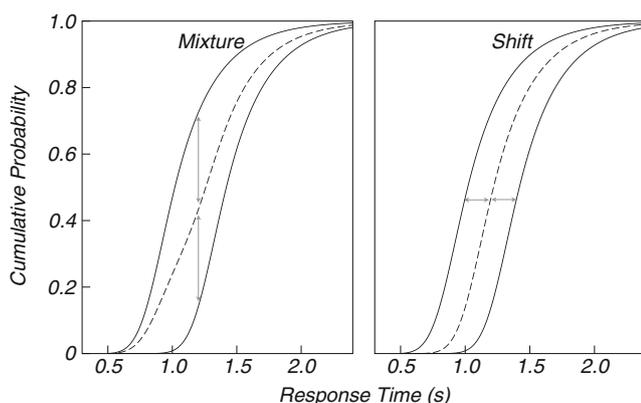


Fig. 1 Depiction of the predictions of the mixture and shift models of the target distribution function. On the left, the distribution of the target is an average of the distributions of the fast and slow reference conditions; that is, each point on the dashed line is a vertical average of the two solid lines. On the right, the distribution of the target is a quantile average of the reference distributions; that is, each point on the dashed line is a horizontal average of the distribution functions

the target distribution that can be expressed in terms of quantiles (such as the interquartile range) will be a weighted average of the corresponding properties of the reference distributions. The quantile average is thus a generalization of the idea that the target is a shift between two distributions that differ in location, and I will refer to the quantile average in this context as a (generalized) shift to highlight its connection to the default assumption about the effects of an experimental factor. A quantile average can be thought of as a horizontal average of the distribution functions, as is shown in Fig. 1.

The multinomial likelihood index describes the evidence for a mixture relative to a (generalized) shift. It does so by using the quantiles of the reference distributions to estimate quantiles of the target distribution under the assumption that the target is either a mixture or a shift. Similar to the logic of Yantis et al. (1991), the estimated target quantiles determine a set of interquantile bins, and the probability of an observation falling into one of these bins follows a multinomial distribution. When the quantiles are equally spaced, the multinomial bins are equally likely with probability $1/b$, where b is the number of bins. In particular, for a set of bin counts n_i , $N = \sum n_i$, the likelihood is:

$$\begin{aligned} \Pr(n_1, n_2, \dots, n_b; p_1, p_2, \dots, p_b) &= \binom{N}{n_1 n_2 \dots n_b} p_1^{n_1} p_2^{n_2} \dots p_b^{n_b} \\ &= \binom{N}{n_1 n_2 \dots n_b} \left(\frac{1}{b}\right)^N. \end{aligned}$$

For computational convenience, I use the log likelihood, and because estimates of n_i are generally fractional, the gamma function is used instead of a factorial. The log likelihood is thus:

$$\ln \Pr(n_1, n_2, \dots, n_b; b) = \ln \Gamma(N + 1) - \sum \ln \Gamma(n_i + 1) - N \ln b. \quad (4)$$

In order to quantify the evidence for a mixture relative to a shift, one calculates the (multinomial) likelihood of the data assuming a mixture and compares it to the likelihood of the data assuming a shift. For both the mixture model and the shift model, the value of k that maximizes the (log) likelihood is estimated. A much higher maximum likelihood for a mixture would provide strong evidence that the mixture assumption provides a better description of the data.

Two technical details need to be addressed in implementing this idea. First, quantiles for the target distribution need to be estimated given a particular value of k . For the shift model, estimating target quantiles is simple: Each target quantile is a weighted average of the quantiles for the reference distributions, and these can be estimated from Eq. 3 for any value of k . Thus, for a given set of quantile probabilities, one can find the corresponding quantiles in the reference distribution and estimate those quantiles for the target distribution. That is, if $R_f^{(i)}$ and $R_s^{(i)}$ are the quantiles

for the fast and slow reference conditions corresponding to probability p_i , then the quantile estimates for the shift model are:

$$\hat{F}_{shift}^{(i)} = kR_f^{(i)} + (1 - k)R_s^{(i)}.$$

However, to estimate target quantiles for the mixture model, one needs the entire distribution function predicted by the model for a particular value of k . Consequently, I used the following procedure. First, approximations of the distribution functions for the reference conditions were estimated from the quantiles by linear interpolation. These estimated distribution functions, $\hat{R}_f(t)$ and $\hat{R}_s(t)$, are piecewise linear with inflection points at the reference distribution quantiles, $R_f^{(i)}$ and $R_s^{(i)}$, respectively. Then for a given value of k , the mixture model provides an estimate of the distribution function for the target based on Eq. 2:

$$\hat{F}_{mixture}(t) = k\hat{R}_f(t) + (1 - k)\hat{R}_s(t).$$

This predicted function is piecewise linear with inflection points at $\{R_f^{(i)}, R_s^{(i)}\}$, the union of the inflection points for the estimated reference distributions. Because it is piecewise linear, it can easily be inverted to produce an estimate of the inverse distribution function for the target. In turn, this can be used to identify the target quantiles at the points p_i :

$$\hat{F}_{mixture}^{(i)} = \hat{F}_{mixture}^{-1}(p_i).$$

Critically, the quantiles are estimated for the same set of probabilities in both the mixture and the shift models, and, for both models, there is only a single free parameter, k .

The second technical issue concerns how bins are estimated for the ends of the distributions. Estimates of the 0 and 1 quantiles in a sample of data consist of the smallest and largest observations, respectively, but these do not provide particularly stable estimates of the population values. Moreover, one could plausibly assume that the distributions are theoretically unbounded at the upper end and bounded at the lower end only by zero. For these reasons, the predicted bins for the target data were assumed to be unbounded at the upper and lower ends, and extreme data were included in those bins. This is tantamount to defining the upper and lower bins as “in this bin or higher” and “in this bin or lower,” respectively. However, the estimated quantiles for the mixture model do depend to some extent on quantile estimates for 0 and 1, because these determine the slopes of the bottom and top piecewise linear approximations to the distribution function.

Bin counts for the observed data were estimated from the target data quantiles rather than the actual observations. As will be discussed below, this makes it easy to apply the present technique to data aggregated across subjects. In order to estimate bin counts for the set of bins predicted

by either the mixture or the shift model, I assumed that the data were evenly distributed across the observed target interquantile bins. This is tantamount to assuming that the piecewise linear approximation to the distribution function for the target condition, \hat{F} , is accurate. The total number of observations in those bins was then apportioned to the bins predicted by the shift or mixture models on the basis of the overlap of the predicted and the observed bins. Thus, in effect, the calculated likelihood provided an index of how well the observed matched the predicted quantiles. This can be written as

$$\hat{n}_i = N \left[\hat{F}(\hat{F}^{(i+1)}) - \hat{F}(\hat{F}^{(i)}) \right],$$

where \hat{F} is the approximation of the target distribution function derived from the target quantiles, $\hat{F}^{(i)}$ is the i th quantile estimated using either the target or the shift model (i.e., the lower boundary of the predicted i th bin), and N is the total number of observations. Using these values of n_i in Eq. 4 yields the log likelihood of the data for a given value of k for the mixture or the shift model. Because all of the estimates involve only linear interpolation based on the quantile estimates in each condition, calculation of the log likelihoods is computationally simple, and it is easy to search for a value of k that maximizes the likelihoods for both models.

This algorithm for calculating the maximum log likelihood under the two models was implemented in R (R Development Core Team, 2011). Quantiles were estimated using the R function “quantile” and the “Type 8” procedure (which produces approximately median-unbiased estimates; Hyndman & Fan, 1996). Linear interpolation and inversion of the linear functions were accomplished using the function “approx.” Estimates of k were found using the R search function “optimize,” which uses a combination of golden section search and successive parabolic interpolation (based on the function “localmin” in Brent, 1973). The R functions for calculating the multinomial likelihood index are available as supplementary materials.

In order to compare the mixture and generalized-shift models, I calculated the difference in the log likelihoods. Because both models have only one free parameter (k) and because the predictions in both cases are derived using similar calculations from the same set of reference distribution quantiles, it is plausible to suppose that the two models have similar complexity and flexibility in matching data. Assuming that this is true, the difference in log likelihoods provides a measure of the relative quality of the two models that is largely independent of any particular theoretical approach to model selection. For example, this statistic is the same as half the difference in Akaike information criterion values for the two models, an approach that might be

used from an information-theoretic perspective (e.g., Burnham & Anderson, 2010). The difference in log likelihoods is similarly half the difference in Bayesian information criterion values, and, from a Bayesian perspective, could be used to calculate an estimate of the posterior odds of the two models (e.g., Wagenmakers, 2007). As a rough guide, differences in the log likelihood of less than 2 would provide a minimum of evidence for one model over the other; values greater than 2 would constitute progressively stronger evidence for one model or the other.

The multinomial likelihood index is based on several assumptions that constrain its theoretical interpretation. First, given either the mixture or the shift model, the quantiles of the target distribution are derived from the empirically observed reference distributions under the assumption that these distributions are fixed and given. Thus, the value of the index should be understood as the likelihood of the target data being a (mixture or shift) combination of those empirical distributions (rather than the population distribution from which the reference data are sampled). Second, as described above, the calculation of the target quantiles uses linear interpolation to provide ad hoc approximations of the distributions. This approach introduces additional noise into those quantile estimates. Thus, although the multinomial likelihood index can be regarded as an estimate of the likelihood given either a mixture or a shift model, the distribution of the index and its relationship to the theoretical likelihood of the data are unknown. However, this is a general feature of likelihood procedures that involve the estimation of parameters. My solution to this issue in the present context is to use Monte Carlo simulations to examine empirically the performance of the index under a range of circumstances that may be typical of response time experiments. If the index allows one to discriminate the two models in a reliable and unbiased fashion, it would be useful despite these approximations.

Effectiveness of the multinomial index

In this section, I describe a variety of Monte Carlo simulations designed to assess how well the multinomial likelihood index technique discriminates mixture and shift distributions under a range of conditions. The base configuration consists of two reference distributions selected to be not unlike those in a response time studies with modest power. The ex-Gaussian distribution was used as the base form, with parameters $\mu = 0.80$ s, $\sigma = 0.12$ s, and $\tau = 0.30$ s; the standard deviation for this distribution is $\sqrt{\sigma^2 + \tau^2} = 0.32$ s. When not otherwise specified, the simulations consisted of 500 observations in the reference and target distributions, and quantiles were estimated for $p = 0, .1, .2, \dots, 1.0$, determining ten bins.

Effect of spread The most potent factor determining the ability to distinguish a mixture from a shift is the spread between the reference distributions. When they are close, the mixture and shift models predict very similar distributions because both are constrained to lie between the fast and slow references. When the reference distributions are sufficiently far apart, the two models are readily distinguished by any number of means. For example, the variance of a mixture will be substantially larger than that of a shift, and a visual inspection of the target distribution may reveal an obvious bimodality. In the simulation, I varied the spread by increasing the mean of the slow reference distribution from 0.85 to 1.20 s in steps of 0.05 s. The mix/shift parameter, k , was fixed at .5, and the mean of the fast reference distribution was fixed at 0.80 s. There were 500 observations in each condition, $\sigma = 0.12$ s, and $\tau = 0.30$ s. The results are shown in Fig. 2, plotted in terms of d , the ratio of the spread to the standard deviation. In this figure (and many of the following ones), the distribution of the difference in log likelihoods in favor of the shift model across 100 simulations is shown as a box plot. Box whiskers are used to indicate the .025 and .975 quantiles. The open box plots show the results when the simulated data did in fact come from a mixture, while the shaded box plots show the results obtained with simulated shift data. A large positive value of the difference in log likelihoods indicates that the mixture model provides a better account of the data; a large negative value indicates that the shift model provides a better account. Values between -2 and 2 would not be particularly strong evidence for either model and might be regarded as ambiguous. This region is shown in gray in the figure. Thus,

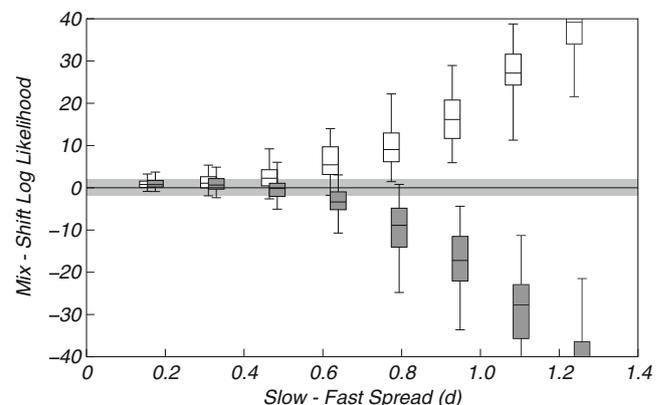


Fig. 2 Effect of spread between the fast and slow reference distributions as a function of d , the ratio of the differences of the means from the standard deviation. Each box plot depicts the median and interquartile range; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

the index would be misleading to the extent that the tail of the upper box plot extends below this region in the negative direction or that the tail of the lower plot extends above the region in the positive direction. The test is effective to the extent that the open box plot is above this region and the shaded box plot is below.

As can be seen from the figure, with the moderate sample sizes used in this simulation, the index is reasonably effective as long as the separation between the reference distributions is larger than about $d = 0.5$. Generally, when the separation is smaller, the index fails to provide clear evidence in favor of either model. Regardless of the separation, though, the index suggests an incorrect conclusion only infrequently. Furthermore, there is relatively little systematic bias: The index does not strongly favor mixtures over shifts or vice versa.

Effects of sample size Generally, one may expect that the power to discriminate the two possible models should increase with sample size. In the present simulations, I explored the effect of increasing sample size for a situation in which a minimum of evidence might be expected with a moderate sample. In particular, the means of the fast and slow reference distributions were set to 0.8 and 1.0 s, respectively, corresponding to $d = 0.62$. The sample size started at 500 and was increased by 0.15 log units at each step. The results are shown in Fig. 3. Although increasing sample size does improve the effectiveness of the index, substantial improvement is only found with an increase to 2,000 or more.

Effects of mixing/shift proportion In the previous simulations, I assumed that the mixing or shift proportion, k , was

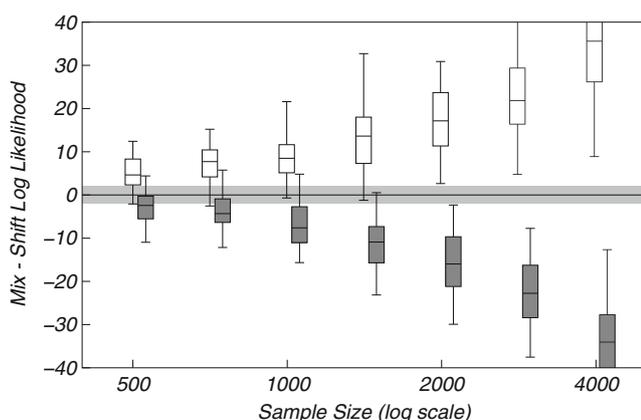


Fig. 3 Effect of sample size. Each box plot depicts the median and interquartile range; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

.5—that is, that the target distribution was halfway between the fast and slow reference distributions. This is clearly the ideal situation for distinguishing the two models. Intuitively, for example, if the value of k was near 0 or 1, the target distribution should be close to the fast or slow reference distribution, regardless of whether it was a mixture or a shift, and the two models would be difficult to disentangle. In order to see how the multinomial likelihood index fares with different values of k , I set the mean of the slow reference distribution to 1.1 s (with the mean of the fast reference distribution set at the default value of 0.8 s) to produce a relatively effective test with $k = .5$, and then varied k from .1 to .9 in steps of .1. As before, $N = 500$, $\sigma = 0.12$ s, and $\tau = 0.30$ s. The results are shown in Fig. 4.

Somewhat surprisingly, k has little effect on the ability of the index to identify a mixture distribution, but small or large values of k tend to produce ambiguous results with data from the shift model. This asymmetry is due to the unique property of mixture distributions (described, e.g., by Sternberg, 1973) that they contain some of the fastest and slowest observations from the reference distributions. Thus, for example, if k is .9, the target will contain mostly observations from the fast reference, but there will be at least a few from the slow reference, and these will produce a distinctive long tail that a shift model would not predict. On the other hand, if the target is a small shift from the fast reference to the slow reference, this can be mimicked without too much difficulty by a mixture model with an even smaller value of k that would minimize the likelihood of having any observations from the slow reference distributions.

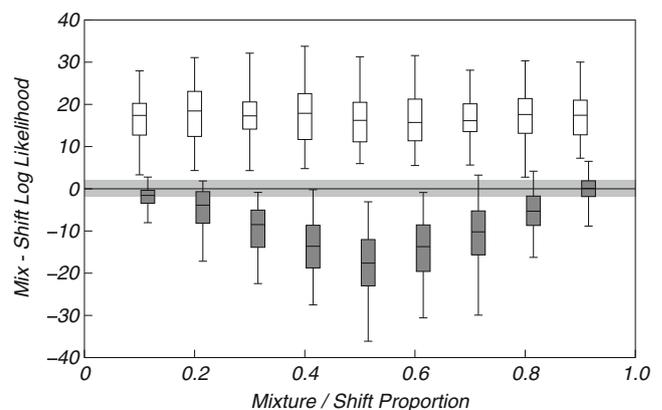


Fig. 4 Effect of the mixture/shift proportion parameter, k . Each box plot depicts the median and interquartile range; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

Selecting the number of bins The number of bins used to calculate the multinomial likelihoods represents a trade-off between estimating the likelihood of an observation falling into each bin and identifying the distinctive shapes of shift and mixture distributions. This critical information concerning the distribution shape is often found in the tails. Decreasing the number of bins increases the number of observations in each bin, and thus improves the estimate of bin probability. However, increasing the number of bins makes it easier to identify distinctive characteristics of the target distribution shape. (Note that this trade-off is a general property of binning the data and also applies, e.g., to the procedure of Yantis et al., 1991.) In principle, the optimal number of bins could vary with the spread of the reference distributions, with mixing/shift proportions, and with the distribution shape. Here, I touch on only one feature of the data, sample size. I performed one series of simulations with a sample size of 500 and another with a sample size of 2,000, spanning what one might think of as the range of small to moderately large studies. In both cases, the mean of the fast reference was set to 0.8 s, and the mean of the slow reference distribution was set to 1.0 s, for a relatively small spread.

The results for the small sample (Fig. 5) suggest that at least eight bins are needed to be effective, and that there is little improvement in the effectiveness of the index with more than 12 bins. With a larger sample (Fig. 6), the index is reasonably effective with 12 bins as well, although some improvement is found with as many as 16 bins. Generally, I suspect that 10–12 bins is likely to be sufficient to capture most of the critical information about distribution shape, given the distribution form assumed here, even with large samples.

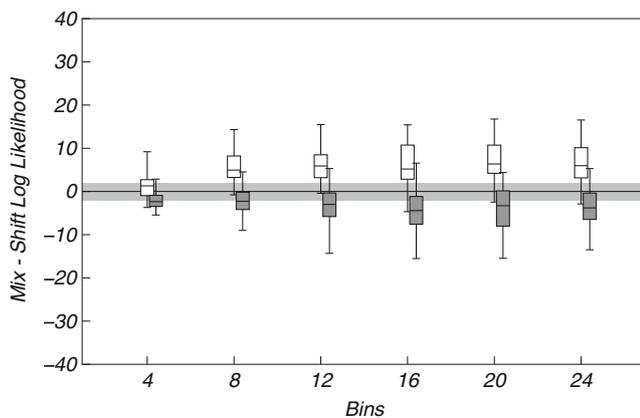


Fig. 5 Effect of the number of bins, with a sample size of 500. Each box plot depicts the median and interquartile range; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

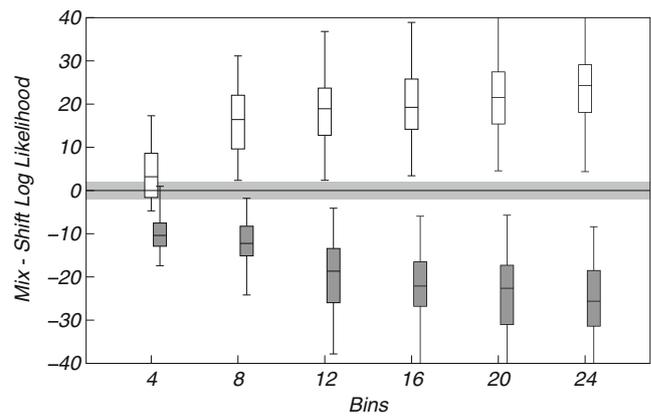


Fig. 6 Effect of the number of bins with a sample size of 2,000. Each box plot depicts the median and interquartile range; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

Data trimming In the simulations presented so far, quantiles were estimated for the entire range of probabilities, from 0 to 1. The 0 quantile is the minimum of the data set, and the 1 quantile is the maximum. Such estimates are unstable, and one might worry that using such quantiles might adversely affect the effectiveness of the multinomial index. As an alternative approach, one may use quantiles in the range $[\varepsilon, 1 - \varepsilon]$, where ε is some suitably small value, and then disregard any data beyond those values. I considered values of ε that were some proportion, α , of the bin probability. For example, if $\alpha = 1$, this would amount to trimming one bin from either end of the distribution, and if $\alpha = .5$, it would amount to trimming half a bin from either end. If there are b bins, the bin boundaries would be at quantiles for the probabilities,

$$\frac{\alpha}{b + 2\alpha}, \frac{1 + \alpha}{b + 2\alpha}, \frac{2 + \alpha}{b + 2\alpha}, \dots, \frac{b + \alpha}{b + 2\alpha}.$$

In order to assess the effect of this form of trimming, I performed a simulation using a mean of 0.8 s for the fast reference and 1.0 s for the slow reference distribution, and trimmed the data at each end of the distribution by various proportions, α , ranging from 0 to 1. As before, $N = 500$, $\sigma = 0.12$ s, and $\tau = 0.30$ s. The results are shown in Fig. 7. As can be seen, with these simulated data distributions, trimming the data has relatively little effect up to about .5 of the bin probability. Trimming more than that causes the results to be more variable and introduces a bias in favor of the shift model. This is perhaps surprising, given the variability of the estimates for the 0 and 1 quantiles. However, critical information for identifying a mixture distribution is found in

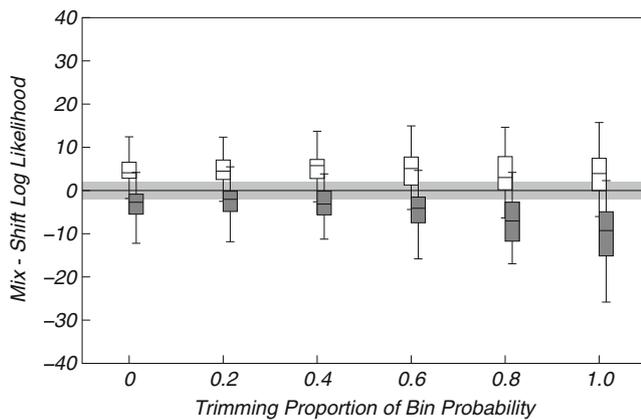


Fig. 7 Effect of data trimming as a function of the proportion of bin probability removed from each end of the distribution. Each box plot depicts the median and interquartile range; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

the tails. Indeed, a general property of using the multinomial distribution to distinguish mixtures and shifts is a trade-off between the stability of the bin estimates and the model-specific information in the distribution tails. The result is that, for moderate amounts of trimming, discarding this information has an adverse effect on the index that offsets any increase in the stability of the quantile estimates. With more extensive trimming, more of the diagnostic information for mixture distributions is lost, resulting in a bias toward the shift model. It is important to note that actual (rather than simulated) data might have true outliers that do not occur in the present simulations, and a small amount of trimming might be useful in such cases. However, the present simulation suggests that trimming more than half a bin probability on either end might reduce the effectiveness of the index.

Aggregating over subjects One of the advantages of constructing an index purely in terms of quantiles is that there is a simple approach to aggregating data from multiple subjects. As suggested by Ratcliff (1979), one may derive an aggregate description of a distribution shape by averaging the quantiles calculated separately for each subject. Such a procedure preserves information about distribution shape even when data from different subjects differ in scale and location. Although averaging quantiles can distort the pattern of results under some circumstances (cf. Van Zandt, 2000), it works reasonably well under a broad range of subject variation. Here, I explore the use of quantile-averaged data with the multinomial likelihood index.

In order to produce a range of different distribution shapes to simulate randomly selected subjects, I assumed

the same mean value of the distribution parameters as in the previous simulations, but allowed each subject to have parameter values that were randomly selected from a uniform distribution centered on that mean. In particular, μ was sampled from the range 0.35 to 1.15 s; σ was sampled from 0.09 to 0.15 s; and τ was sampled from 0.2 to 0.4 s. The density for a random sample of six subjects generated using these parameter ranges is shown in Fig. 8. Although these distributions may not capture the kinds of intersubject variation seen in any particular study, they at least provide a range of variation across subjects that will allow an initial assessment of whether such variation matters with the present technique.

In the first simulation, I examined the effect of number of subjects. In this case, I assumed that each subject received 50 trials and that the mean of μ_f across subjects was 0.8 s and the mean of μ_s was 1.0 s. Thus, this simulation is a replication of the simulation of sample size in Fig. 3, except that the number of (randomly selected) subjects was varied instead of the size of a single sample. The number of subjects in the smallest simulation was 10, and this was increased by 0.15 log units at each step. The results are shown in Fig. 9. As can be seen, the results are quite similar to those obtained with a single homogeneous sample. Thus, given the range of subject variation considered here, it would seem that the critical determinant of effectiveness is the total number of observations, not the number of subjects or the number of observations per subject.

One unrealistic aspect of the previous simulation is that, for each subject, the fast and slow reference distributions were assumed to have the same (randomly selected) shape. It is quite reasonable to suppose that those two conditions will vary in other ways than location, even within a subject. To assess the extent to which such variation affects the results, I selected the parameters for the fast reference distribution from the same ranges as I did previously. I then performed two sets of simulations. In the first, I selected a

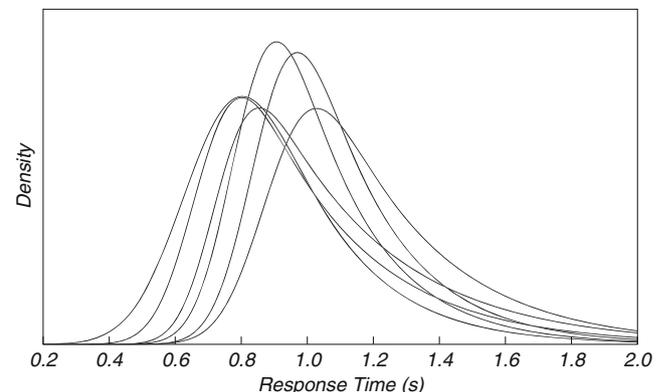


Fig. 8 Distributions from six randomly generated subject simulations. See the text for the parameter sampling details

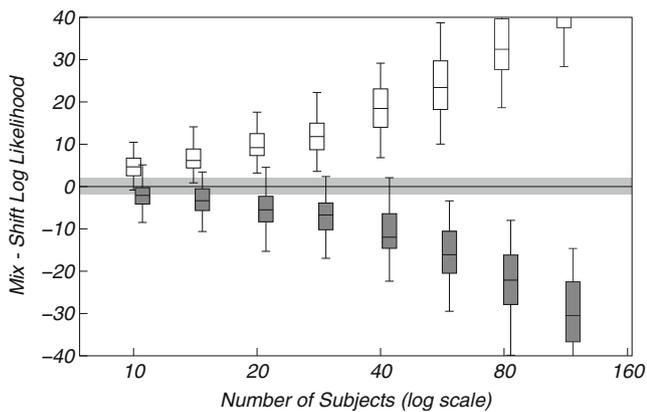


Fig. 9 Effect of number of subjects, each with 50 observations. Each box plot depicts the median and interquartile range; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

range for the parameters of the slow reference distribution relative to the corresponding parameters of the fast reference distribution. This range was as follows: μ was 0.2 s larger than the faster slow reference, ± 0.2 s; σ was varied by ± 0.03 s; and τ was varied by ± 0.1 s. The slow reference parameters were sampled from a proportion of those ranges in different simulations, with the proportion ranging from 0 (implying no difference between the fast and slow reference parameters) to 1 (the entire range). In the second set of simulations, I assessed the effect of variations in k by selecting randomly for each subject. The parameter was sampled from some proportion of the range 0 – 1 . That proportion ranged from 0 to 1 in different simulations. When the proportion was 0 , the parameter was fixed at $.5$; when the proportion was 1 , the parameter was sampled from the full range of 0 – 1 . In both sets of simulations, the number of subjects was set at 40 , so that a reasonably effective comparison would be obtained when the variation was 0 . The results are shown in Figs. 10 and 11.

As can be seen, there is little effect of variations in the parameters. The general conclusion suggested by these simulations of aggregate data is that as long as the distributions have the same general form for all subjects, the index can be applied without concern to an aggregate data set generated by quantile averaging. In particular, the effectiveness of the index increases with the number of subjects, just as it does with total N in the single-sample simulations; variations in distribution shape across conditions have relatively little effect; and variations in the shift/mixture proportion across subjects have little impact. I have operationalized the notion of “same general form” as having ex-Gaussian distributions with parameters that produce plausible descriptions of

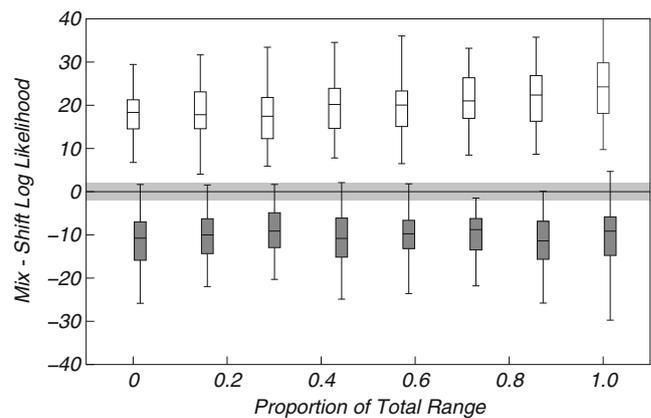


Fig. 10 Effect of variation across conditions, expressed as proportions of the total range; see the text for the distribution of parameter values. Each box plot depicts the median and interquartile ranges; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

response time distributions. However, I suspect that the use of the ex-Gaussian here is not critical and that my conclusions would hold with other distribution shapes that might be found with response times, as long as they are unimodal with modest skew. Of course, with more disparate distribution shapes across subjects and conditions, it is quite possible that quantile averaging would not provide a reasonable characterization of the data set as a whole, and the use of the index on such aggregated data would not be expected to produce sensible results.

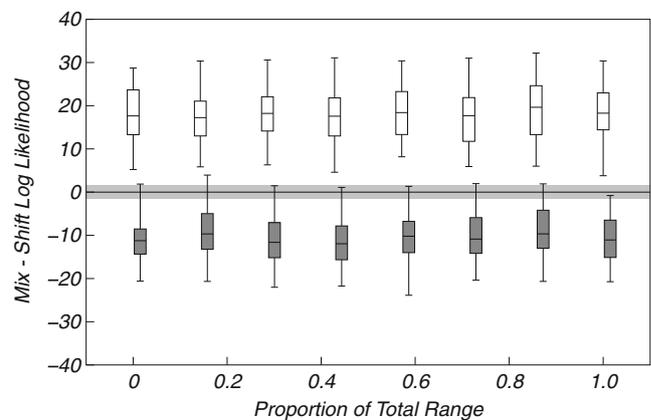


Fig. 11 Effect of variation in the mixing/shift parameter k , expressed as a proportion of the total range, 0 – 1 . Each box plot depicts the median and interquartile range; whiskers indicate the .025 and .975 quantiles. Increasingly positive results indicate increasingly strong evidence for a mixture, and the gray region between -2 and 2 indicates results that might be ambiguous. The open boxes depict results with simulated mixtures, and the gray boxes depict results with simulated shifts. The boxes for simulated shifts are plotted slightly to the right for clarity

Applications of multinomial likelihood index

In this section, I apply the multinomial index to several existing data sets to demonstrate its effectiveness with real data. First, I apply it to data that might reasonably be expected to conform to a distribution mixture—namely, data from investigations of residual switch costs in the task-switching paradigm. Second, I apply the index to a paradigm for which the common theoretical interpretation is a shift, the psychological refractory period paradigm. Finally, I apply the index to a situation in which a theoretical interpretation in terms of either a shift or a mixture might be possible, the variation in the Simon effect with trial sequence. The application of the index in all of these examples makes the simplifying assumption that there are only two possible distributions. However, as discussed above, the reference distributions need not be pure examples of those distributions, and might be mixtures themselves. Possible violations of this “two-distribution” assumption are discussed in the concluding paragraphs.

Residual switch cost In the task-switching paradigm, subjects perform two or more tasks on different trials. The common finding is a “switch cost”: Response times are slower if the task performed on the current trial is different than that performed on the previous trial (e.g., Altmann & Gray, 2008; Jersild, 1927; Rogers & Monsell, 1995). A common interpretation of switch costs is that some set of cognitive operations are needed to prepare for the new task and that these operations take time. Monsell (2003) and others refer to these operations as “task reconfiguration.” (However, for alternative perspectives on switch costs, see, e.g., Allport, Styles, & Hsieh, 1994; Waszak, Hommel, & Allport, 2003; and Logan & Bundesen, 2003.) From this theoretical perspective, one would predict that switch costs should decrease when the task change is predictable and subjects have sufficient time to accomplish the reconfiguration operations. Indeed, under many circumstances, switch costs *do* decrease with the amount of time that subjects have to prepare for a task switch (see Rogers & Monsell, 1995). However, a number of authors have found that switch costs remain substantial even at relatively long intervals and even when subjects might be expected to be motivated to completely prepare on each trial (e.g., Nieuwenhuis & Monsell, 2002). This effect has been referred to as “residual switch cost.”

De Jong (2000) proposed a mixture model of residual switch costs. He argued that when there should be ample time to prepare for a task switch, subjects “fail to engage” with the task on some proportion of those trials. According to this logic, response times on long-delay switch trials should be a mixture of trials on which subjects are prepared for the new task and trials on which they are unprepared.

Furthermore, the distribution of prepared trials should resemble that for the long-delay repeat trials, while the distribution of unprepared trials should resemble that for short-delay switch trials (where there is no question that subjects have insufficient time to prepare for the new task). De Jong tested the fit of the mixture model using the multinomial goodness-of-fit test described by Yantis et al. (1991), failed to reject the null hypothesis, and concluded that the mixture model provided a good account of residual switch costs. Others have been less sanguine about this interpretation. Lien, Ruthruff, Remington, and Johnston (2005) argued that subjects may have been differentially prepared for different stimuli rather than being in different preparatory states across trials. Brown et al. (2006), using a powerful experimental design with nearly 2,000 observations per condition, rejected the null hypothesis of a mixture distribution using a version of the fixed-point property.

In order to explore the effectiveness of the multinomial likelihood index, I applied the technique to some of the results from Brown et al. (2006). In that study, subjects performed two trials in a row of a letter task (in which a stimulus is classified as a vowel or a consonant), followed by two trials of a digit task (in which a stimulus is classified as odd or even). Thus, the first trial of each task sequence was a switch trial, while the second was a repeat trial. I focus here on the data from the letter task (although the data from the digit task were similar). The distributions for that task in the long- and short-delay switch trials and the short-delay repeat trials were estimated from Brown et al.’s Fig. 3 and are plotted in Fig. 12. The long-delay condition in this experiment involved a 1.20-s delay between the response on one trial and the stimulus for the next; the short delay was 0.15 s. Because the Brown et al. figure provides quantiles from .05 to .95 at .05 intervals, I used these to define the bins, excluding the top and bottom 5 % of the data. Applying the index provides a log likelihood of -85.10 for the mixture model and -132.35 for the shift model, or a difference of

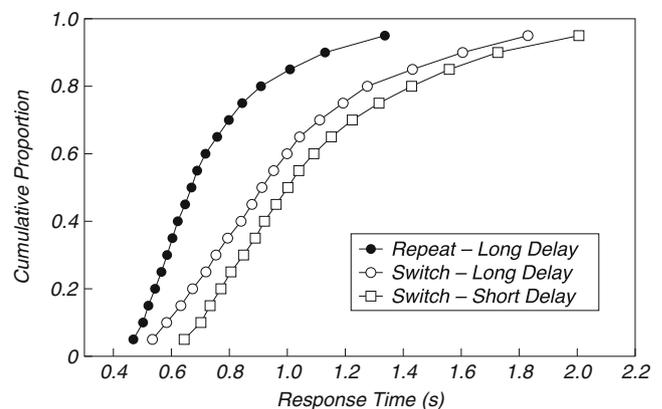


Fig. 12 Cumulative distribution functions for the letter condition in Brown et al. (2006)

47.25 in favor of the mixture model. The estimated values of k were .20 for the mixture model and .37 for the shift model. Clearly, the data overwhelmingly support the mixture model. This is apparent in Fig. 12, in which the target (long-delay switch) condition has a somewhat larger variance and spreads into the left tail of the fast reference distribution and the right tail of the slow reference distribution. This pattern is characteristic of a distribution mixture.

The fit of the mixture and shift models to the long-delay switch condition data is shown in Fig. 13. In this figure, the data from the target condition are plotted as a histogram, with the bars corresponding to the equally likely bins used in the model fits. The gray-shaded area corresponds to the actual data, and the outlined histogram indicates the fit of the model, with the mixture model on the left and the shift on the right. Although the fit of the mixture model deviates to some extent from the actual data in the tails, it provides a much closer match than the overall shift model.

The results from the multinomial likelihood index support the mixture model over the shift model, while Brown et al. (2006) rejected the mixture model account using the same data. There are several possible interpretations of the discrepancy between the conclusion suggested by the multinomial likelihood index and that reached by Brown et al. First, the present index provides a comparison between two models and identifies which of the two provides a better account. In contrast, Brown et al. performed a significance test, with the mixture model as the null hypothesis. Moreover, because of the size of the design, the test is likely to have had a great deal of power. A well-known critique of the hypothesis-testing framework is that in many contexts, the null hypothesis can be assumed to be incorrect in detail, and that a sufficiently powerful test is guaranteed to reject it (e.g., Cohen, 1990). Thus, one might suspect that the Brown test provides an example of statistical but not theoretical significance.

A related but perhaps more insightful interpretation is that although the present index favors the mixture model,

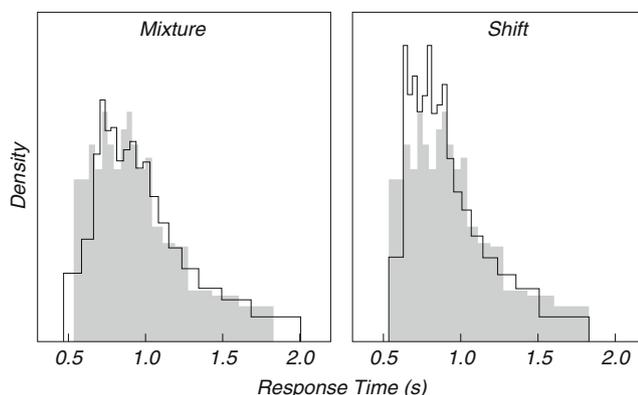


Fig. 13 Predicted and observed interquantile bins for the mixture and shift models when applied to the letter data from Brown et al. (2006)

that model does not provide an overwhelmingly precise match to the results; the index favors the mixture model only because the shift model is even worse. Thus, one possible interpretation is that neither model fits well. Indeed, one could imagine performing a significance test of the null hypothesis that the target distribution is a (generalized) shift between the fast and slow reference. Such a test might be based on whether, for example, the interquartile range in the target distribution is larger than the average of that in the reference distributions. It seems possible that such a significance test would also reject the null hypothesis. From this perspective, one might argue that although the target condition is not precisely a mixture, neither is it a shift. Brown et al. (2006) provided a discussion of their results that is generally consistent with this perspective. Rather than dismissing the mixture model completely, they suggested that the long-delay switch trials may be a more complex mixture than is predicted by a simple form of the failure-to-engage account.

Psychological refractory period As a demonstration that the multinomial likelihood index can provide evidence for a shift as well as a mixture, I applied it to data from the psychological refractory period paradigm. In this situation, subjects are presented with two stimuli separated by a relatively brief SOA and make a response to each in turn. The overwhelmingly common finding in this paradigm is that response times to the second stimulus increase as the SOA is decreased. At short SOAs, the slope of this increase can approach -1 , suggesting that the second response is made a fixed time after the first stimulus is presented. A popular interpretation of this pattern of results is a “bottleneck” model in which some central stage of processing can only be used in one task at a time (Pashler, 1994). Thus, at short SOAs, the second task must wait until the first task completes this central stage before it can be used for the second task. A variety of findings converge on the conclusion that the central stage in question involves selection of the response (e.g., McCann & Johnston, 1992).

Although the bottleneck model provides a parsimonious account of a range of results, it is not without its critics. Meyer and Kieras (1997) proposed an account based on strategic stopping in the context of their EPIC model of complex task performance. In this account, the mental processing of both tasks can, in principle, proceed in parallel. However, subjects halt processing of the second task at certain points to comply with the task demand that the response to the first stimulus must be made prior to the response to the second stimulus. Tombu and Jolicœur (2003) have also argued that the central process in the bottleneck model need not be all-or-none, and that capacity sharing between the two tasks can occur under some circumstances. From the perspective of the present demonstration, though, the important generalization of all of

these theoretical accounts is that responses to the second task are progressively delayed as the SOA is decreased. Thus, the prediction is that medium SOAs should be a shift between short-SOA trials (on which there is a large amount of delay) and long-SOA trials (on which there is relatively little delay). In particular, none of these accounts would predict a mixture distribution.

In order to assess this prediction, I used data collected by Duffels (2010, Exp. 1) in my lab. In the first task, subjects decided whether a brief sound was a “beep” or a “tap” and responded with the left index and middle fingers. In the second task, subjects decided whether two perspective views of a room depicted the same room or different rooms. The rooms were square in layout with a single colored wall and a small ball located in one corner, and they were depicted from a variety of different orientations. Thus, the same/different response required subjects to identify whether the rooms in the two views had the same relationship between the ball and the colored wall. Responses to the second task were made with the index and middle fingers of the right hand. The SOA between the tone and the room views was 0.1, 0.2, 0.3, 0.5, 0.7, or 1.0 s. However, for the present application, I only considered second-task response times from a short-SOA condition (0.1 s), a medium-delay condition (0.3 s), and a long-delay condition (0.7 s). There were 20 subjects, each contributing approximately 96 correct responses to each of the three conditions.

In this application, ten bins were used, and a half-bin probability of data was trimmed from the upper and lower tails. The aggregate data are shown in Fig. 14. As can be seen in this depiction of the data, the medium-SOA condition shows little sign of including observations from the tails of the long- and short-SOA conditions. Although the distribution of times in the long-SOA condition has a somewhat different shape, the principal difference between the distributions for the three conditions seems to be one of location. Thus, the overall appearance of the data suggests a shift

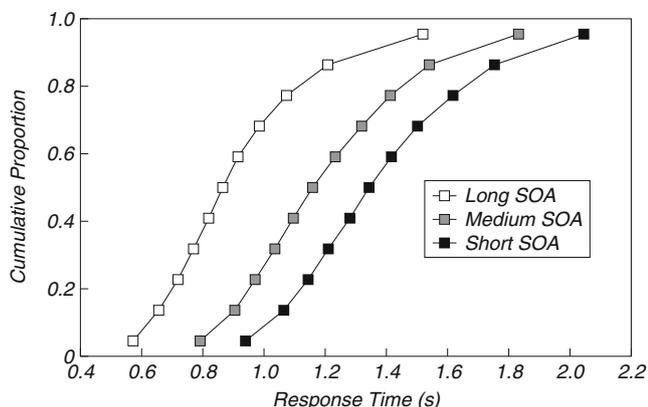


Fig. 14 Cumulative distribution functions for the short-, medium-, and long-SOA conditions from Experiment 1 of Duffels (2010)

rather than a mixture. The multinomial likelihood index confirms this impression. The log likelihood for the shift model was -37.08 , while for the mixture model it was -192.30 . The fit values of k were .39 for the shift model and .16 for the mixture model. The two fits are shown in Fig. 15. Clearly, the shift model provides a close approximation of the obtained data, while the mixture model is inadequate.

Trial sequence interactions with the Simon effect In this final section, I describe an application of the multinomial likelihood index to a situation in which both the shift model and the mixture model are plausible. The Simon effect is a compatibility effect between the (irrelevant) location of a stimulus and the side of the response: Responses are faster when the stimulus is presented on the same side as the response (Lu & Proctor, 1995; Simon, 1969; Simon & Rudell, 1967). Interestingly, the Simon effect is often found to occur with fast responses and not slow responses (e.g., Hommel, 1993). Thus, for example, the fast tail of the response time distribution is faster when the stimulus is compatible with the response, while slower responses are relatively unaffected by compatibility. This in itself suggests a mixture: Perhaps some responses are made quickly on the basis of stimulus location, leading to a mixture of responses made with respect to location and generally slower responses made on the basis of stimulus identify.

The Simon effect, like several other related compatibility effects, interacts with trial sequence. In particular, the effect is much less pronounced following an incompatible trial than following a compatible trial. This pattern of results was first reported by Gratton et al. (1992) using the flanker task and has been investigated by Botvinick, Nystrom, Fissell, Carter, and Cohen (1999) and Verbruggen, Notebaert, Liefoghe, and Vandierendonck (2006), among others. Botvinick, Braver, Barch, Carter, and Cohen (2001) developed a model of cognitive control that explains such effects in terms of a process of conflict monitoring. This

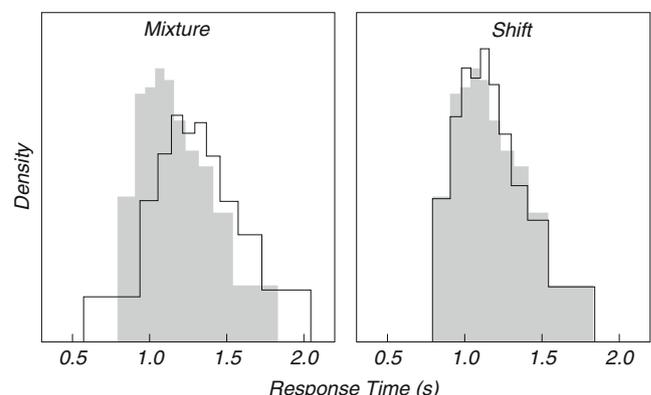


Fig. 15 Predicted and observed interquartile bins for the mixture and shift models when applied to the data from Experiment 1 of Duffels (2010)

sequential dependency has been investigated with the Simon effect by, for example, Stürmer, Leuthold, Soetens, Schröter, and Sommer (2002). Hommel, Proctor, and Vu (2004) argued that the elimination of the Simon effect following an incompatible trial is due to the integration of the stimulus and response from successive trials (see also Wendt, Kluwe, & Peters, 2006).

Data on sequential effects in the Simon task were obtained in an experiment in my lab. The stimuli were X and O presented in 40-point Helvetica font. They were presented either 75 or 300 pixels to the left or the right of fixation; these correspond to approximately 2.7° or 10.7° of visual angle at a typical viewing distance of 50 cm. The stimuli remained in view until subjects made an identification response with either their left or right index finger. Stimuli were presented on an iMac monitor, and responses were collected using the iMac keyboard. Twelve subjects completed nine blocks of 200 trials, divided evenly among the eight possible stimulus position combinations; the first block was regarded as practice. Half of the subjects responded to the X with their left hand and to the O with their right, while the other half had the reverse stimulus–response mapping. Each trial began 1 s after the preceding response, with a longer delay between blocks. Trials were classified as compatible (a left-hand response to a stimulus appearing to the left of fixation or a right-hand response to a stimulus appearing to the right of fixation) or as incompatible. Incorrect responses were discarded. Because the order of trials in a block was random, each subject performed approximately 400 trials with each combination of compatible or incompatible current trial and compatible or incompatible previous trial. Of these four combinations, responses on compatible following compatible trials were the fastest, and responses on incompatible following compatible trials were the slowest. For the present application, I examined, as the target condition, compatible trials following incompatible trials. These are intermediate between the two types of trials following compatible trials, and the multinomial likelihood index was used to assess whether they might be regarded as a mixture of the faster and slower conditions. The distributions of response times from this experiment are shown in Fig. 16. As can be seen, the general pattern is consistent with a mixture, in that fast responses in the compatible preceded by the incompatible condition are as fast as those in compatible preceded by compatible trials, while slow responses are as slow as those in incompatible preceded by compatible trials.

The fits of the mixture and shift models are shown in Fig. 17. The figure indicates that although the mixture model does not precisely match the shape of the fast tail of the distribution, it does much better than the shift model for the central part of the distribution. The multinomial likelihood index confirms this impression. The log likelihood for the mixture model was -40.28 , as compared to -92.27 for

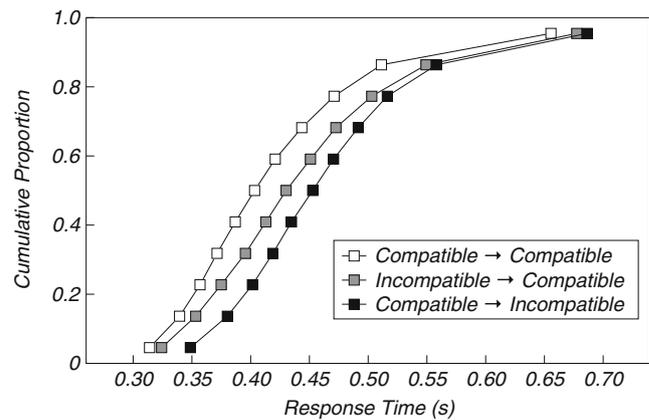


Fig. 16 Cumulative distribution functions in the Simon task for compatible and incompatible trials preceded by a compatible trial and compatible trials preceded by an incompatible trial

the shift model, or a difference of 51.99 in favor of the mixture model. (The estimated values of k were .49 for the mixture model and .55 for the shift model.)

One possible interpretation of a mixture model in this context is based on the idea that the location of the stimulus tends to automatically activate a response on that side (cf. Hommel, 1993). From this perspective, one might hypothesize that there are two types of responses: *facilitated*, in which the automatic activation of the response is consistent with the correct response and facilitates that response, and *inhibited*, in which the automatic activation of a response is suppressed at some cost to response time. In order to explain why compatible trials are slower following an incompatible trial, one could assume that subjects are more cautious following an incompatible trial (cf. Botvinick et al., 2001). As a consequence, subjects may inhibit the tendency to respond on the basis of location on some portion of the trials, even if the stimulus location is compatible. Thus, some percentage of the compatible trials following an incompatible trial would resemble incompatible trials (which, by hypothesis, are always inhibited). Although a variety of

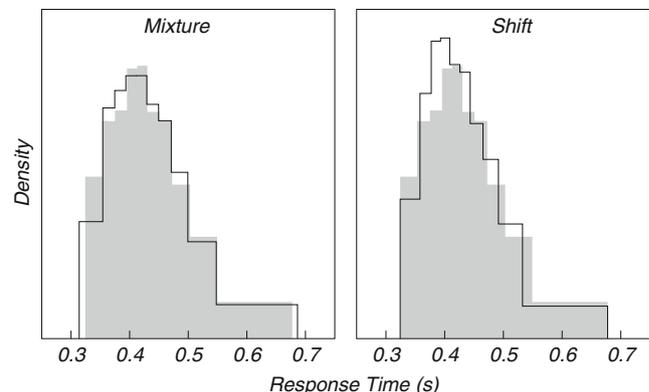


Fig. 17 Predicted and observed interquartile bins for the mixture and shift models when applied to the Simon task data

other aspects of the data from this experiment would need to be considered in order to build a complete account of the results, the relatively accurate fit of the mixture model to these data raises some important theoretical questions.

Distributional approaches to assessing evidence for mixtures

If one is willing to make assumptions concerning the form of response time distributions, one might pursue a variety of other approaches to assess the evidence for mixtures. For example, a common assumption in the modeling of choice responses is that evidence is gradually accumulated over time. Models in this general class include Ratcliff's diffusion model (e.g., Ratcliff, 1988), the leaky-accumulator model of Usher and McClelland (2001), and the linear ballistic accumulator model of Brown and Heathcote (2008). These approaches generally provide a description of response time distributions that is reasonably accurate in a range of paradigms. Importantly, the models provide predictions not only for response times on correct trials, but also for the probability of an error and for the distribution of response times on incorrect trials. There are also several ad hoc, descriptive accounts of response time distributions; these include the ex-Gaussian distribution (used in the simulations presented here) and the shifted-Wald distribution. Both provide reasonably accurate descriptions of the shapes of response time distributions. Furthermore, a number of authors have identified the parameters of these distributions with various mental processes (e.g., Hohle, 1965; Luce, 1986). Any of these approaches could be used to characterize the distributions the shapes of the reference distributions with relatively few parameters. If one proceeded by analogy to the technique developed here, one would then make predictions concerning the shape of the target distribution under the assumptions that the target condition constitutes a shift or a mixture.

A variety of issues arise in pursuing this method. First, fitting these models to the data can be complex and may require large sets of data to provide stable estimates of the parameters (Wagenmakers, 2009; cf. Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009). However, recent advances in fitting methods make this more practical (e.g., Donkin, Averell, Brown, & Heathcote, 2009; Farrell & Ludwig, 2008). Second, one would need to decide whether and how to aggregate across subjects. In some previous applications of these models, models have been fit to individual subjects (e.g., Dutilh et al., 2009; Ratcliff, 2006). Needless to say, this approach requires reasonably large data sets for each subject. Alternatively, it may be possible to average the quantile estimated for each subject (such as was done here) and then to use the average quantiles to estimate parameters. Still another

approach is to use hierarchical models with random subject effects in which one estimates the parameters of the distribution from which the individual-subject parameters are drawn (e.g., Vandekerckhove, Tuerlinckx, & Lee, 2011). Third, one would have to measure the fit of the model to the target distribution. This could be done with, for example, goodness-of-fit tests that assess the adequacy of the mixture and shift models, or a comparison of the fits using a suitable model comparison index (as was advocated here).

Identifying mixtures using these types of theoretical approaches may be more sensitive than the nonparametric approach developed here. However, this additional precision would be at the cost of making strong assumptions concerning the processing involved or the forms of the response time distributions. The extent to which such an approach is robust with respect to violations of these assumptions would need to be assessed. More generally, there may be limited value in making a simple distinction between a description of the data as a shift or a mixture in the context of strong a priori assumptions concerning the forms of the distributions. For example, one might simply construct a full process model of the target condition using the same concepts used for the reference distribution and fit that directly. The best fit of such a parametric model to the target distribution may not be simply a shift or a mixture. A potentially important distinction that might be addressed would be whether the target condition involves a mixture or a shift of the distribution parameters, rather than the distributions themselves. In the context of such a detailed theoretical account, the present approach, which makes a heuristic distinction between two simple models of the target condition, may not be necessary. Rather, the multinomial likelihood index may be more useful in situations in which relatively little is known about the mechanisms that produce response times in the three conditions.

Conclusions

In this article, I have described a technique for assessing the evidence for a mixture model of a response time distribution relative to a (generalized) shift model. In order to use the technique, one needs three experimental conditions: a slow reference condition, a fast reference condition, and a target condition in between the two that could be regarded as either a mixture or a shift. The method is computationally simple and easy to implement, and it can be successfully applied to modest-sized data sets. As demonstrated in the Monte Carlo simulations, the technique is effective at discriminating the two models with even a small degree of separation between the reference distributions. Moreover, the technique is nonparametric and robust with respect to variations in distributions and variations across subjects in the shapes of the distributions and the magnitudes of the

effects. This power derives primarily from comparing two distinct models, rather than performing a significance test or goodness-of-fit test. When applied to several actual data sets, the index generated the expected result in two situations in which theory predicts a mixture and a shift, respectively, and produced a potentially insightful result in a context in which there was no strong theoretical expectation.

The main limitation of the index is also the feature that makes it powerful: It provides a comparison between two simple cases. When the reality is more complex, the index will not necessarily provide a useful contrast, and in some cases, the results could be misleading. Nevertheless, it seems more profitable to theorize about the nature of the data with information gleaned from the index in hand rather than to proceed in ignorance. Moreover, inferences about the means of conditions are quite commonly done without considering the natures of the distributions at all. On the basis of the typical assumptions of the general linear model, it is often implicitly assumed that effects consist of simple changes in location. Thus, a large majority of research using response time measurements has assumed one of the simple cases considered by the present index. Contrasting this simple case with another simple case will almost always provide useful information.

More generally, though, one may argue that the target distribution will virtually never simply be a mixture of or a shift between the two reference distributions. Even if the target distribution is a mixture of performance from two states or strategies, the conditions of the reference distributions differ to some extent from those of the target, and they may not precisely represent the states or strategies used in the target condition. Thus, the distributions for the reference conditions must be regarded as only approximations of the distributions that might be mixed in the target condition. For example, in the failure-to-engage account discussed in connection with the task-switching example, there may be mental processes, such as priming or expectation, that are specific to the repetition of a task. These might make the distribution of response times distinct from that observed on switch trials, even when subjects are entirely prepared for the switch. Similarly, even when the effect of a manipulation is primarily to shift the location of a distribution of response times, details of the condition may change the shape of the distribution to some extent, and these changes need not be captured by quantile averaging, as assumed in the multinomial likelihood index. For example, short SOAs in the psychological refractory period paradigm might lead to response-grouping strategies on a small percentage of trials (cf. Borger, 1963) that are not reflected in the distributions of somewhat longer SOAs. Thus, the match to the generalized-shift model need not be perfect, even when the theoretical interpretation is largely correct. In general, the most appropriate interpretation of the results of the index is

to provide a heuristic interpretation of performance in the target condition that can be used to guide a broader theoretical account of performance in the task.

References

- Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421–452). Cambridge: MIT Press.
- Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review*, *115*, 602–639. doi:10.1037/0033-295X.115.3.602
- Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, *33*, 526–542.
- Borger, R. (1963). The refractory period and serial choice-reactions. *Quarterly Journal of Experimental Psychology*, *15*, 1–12.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652. doi:10.1037/0033-295X.108.3.624
- Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*, 179–181. doi:10.1038/46035
- Brent, R. (1973). *Algorithms for minimization without derivatives*. Englewood Cliffs: Prentice-Hall.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178. doi:10.1016/j.cogpsych.2007.12.002
- Brown, S., Lehmann, C., & Poboka, D. (2006). A critical test of the failure-to-engage theory of task switching. *Psychonomic Bulletin & Review*, *13*, 152–159. doi:10.3758/BF03193827
- Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304–1312. doi:10.1037/0003-066X.45.12.1304
- De Jong, R. (2000). An intention-activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 357–376). Cambridge: MIT Press.
- Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavior Research Methods*, *41*, 1095–1110. doi:10.3758/BRM.41.4.1095
- Duffels, B. (2010). *The stages of processing of one's environment*. Unpublished master's thesis, University of Alberta, Edmonton, Canada.
- Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*, 1026–1036. doi:10.3758/16.6.1026
- Eriksen, C. W., & Eriksen, B. A. (1972). Visual backward masking as measured by voice reaction time. *Perception & Psychophysics*, *12*, 5–8. doi:10.3758/BF03212835
- Everitt, B., & Hand, D. J. (1981). *Finite mixture distributions*. New York: Chapman and Hall.
- Falmagne, J.-C. (1965). Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology*, *2*, 77–124.
- Falmagne, J.-C. (1968). Note on a simple fixed-point property of binary mixtures. *British Journal of Mathematical and Statistical Psychology*, *2*, 77–124.

- Falmagne, J.-C., Cohen, S. P., & Dwivedi, A. (1975). Two-choice reactions as an ordered memory scanning process. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 296–344). New York: Academic Press.
- Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, *15*, 1209–1217. doi:10.3758/PBR.15.6.1209
- Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*, 480–506. doi:10.1037/0096-3445.121.4.480
- Green, D. M., & Luce, R. D. (1971). Detection of auditory signals presented at random times. *Perception & Psychophysics*, *9*, 257–268.
- Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*, 382–386. doi:10.1037/h0021740
- Hommel, B. (1993). The relationship between stimulus processing and response selection in the Simon task: Evidence for a temporal overlap. *Psychological Research*, *55*, 280–290.
- Hommel, B., Proctor, R. W., & Vu, K.-P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, *68*, 1–17. doi:10.1007/s00426-003-0132-y
- Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, *50*, 361–365.
- Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *89*, 1–81.
- Jiang, Y., Rouder, J. N., & Speckman, P. L. (2004). A note on the sampling properties of the Vincentizing (quantile averaging) procedure. *Journal of Mathematical Psychology*, *48*, 186–195. doi:10.1016/j.jmp.2004.01.002
- Klemmer, E. T. (1957). Simple reaction time as a function of time uncertainty. *Journal of Experimental Psychology*, *54*, 185–191.
- Lien, M.-C., Ruthruff, E., Remington, R. W., & Johnston, J. C. (2005). On the limits of advance preparation for a task switch: Do people prepare all the task some of the time or some of the task all the time? *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 299–315. doi:10.1037/0096-1523.31.2.299
- Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 575–599. doi:10.1037/0096-1523.29.3.575
- Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, *2*, 174–207. doi:10.3758/BF03210959
- Luce, R. D. (1986). *Response times: Their role in inferring mental organization*. New York: Oxford University Press.
- Lupker, S. J., & Theios, J. (1975). Tests of two classes of models for choice reaction times. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 137–146. doi:10.1037/0096-1523.1.2.137
- Lupker, S. J., & Theios, J. (1977). Further tests of a two-state model for choice reaction times. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 496–504. doi:10.1037/0096-1523.3.3.496
- McCann, R. S., & Johnston, J. C. (1992). Locus of the single-channel bottleneck in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 471–484.
- Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 2. *Accounts of psychological refractory-period phenomena*. *Psychological Review*, *104*, 749–791. doi:10.1037/0033-295X.104.4.749
- Meyer, D. E., Yantis, S., Osman, A., & Smith, J. E. K. (1985). Temporal properties of human information processing: Tests of discrete versus continuous models. *Cognitive Psychology*, *17*, 445–518.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*, 134–140. doi:10.1016/S1364-6613(03)00028-7
- Nieuwenhuis, S., & Monsell, S. (2002). Residual costs in task switching: Testing the failure-to-engage hypothesis. *Psychonomic Bulletin & Review*, *9*, 86–92. doi:10.3758/BF03196259
- Olds, E. S., Cowan, W. B., & Jolicœur, P. (2000). Partial orientation pop-out helps difficult search for orientation. *Perception & Psychophysics*, *62*, 1341–1347. doi:10.3758/BF03212136
- Ollman, R. T. (1966). Fast guesses in choice-reaction time. *Psychonomic Science*, *6*, 155–156.
- Ollman, R. T., & Billington, M. J. (1972). The deadline model for simple reaction times. *Cognitive Psychology*, *3*, 311–336.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*, 220–244. doi:10.1037/0033-2909.116.2.220
- R. Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org/
- Rabbitt, P. M. A. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*, 264–272.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461. doi:10.1037/0033-2909.86.3.446
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review*, *95*, 238–255. doi:10.1037/0033-295X.95.2.238
- Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*, 195–237. doi:10.1016/j.cogpsych.2005.10.002
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology*, *124*, 207–231. doi:10.1037/0096-3445.124.2.207
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309–316. doi:10.1037/0033-2909.105.2.309
- Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society Serial B*, *43*, 97–99.
- Simon, J. R. (1969). Reaction toward the source of stimulation. *Journal of Experimental Psychology*, *81*, 174–176.
- Simon, J. R., & Rudell, A. P. (1967). Auditory S–R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*, 300–304.
- Sternberg, S. (1973, November). *Evidence against self-terminating memory search from properties of reaction-time distributions*. Paper presented at the 14th Annual Meeting of the Psychonomic Society, St. Louis, MO.
- Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1345–1363. doi:10.1037/0096-1523.28.6.1345
- Thomas, E. A. C. (1969). Distribution free tests for mixed probability distributions. *Biometrika*, *56*, 475–484.
- Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 3–18. doi:10.1037/0096-1523.29.1.3
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592. doi:10.1037/0033-295X.111.3.757
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*, 424–465. doi:10.3758/BF03214357

- Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods, 16*, 44–62. doi:10.1037/a0021765
- Verbruggen, F., Notebaert, W., Liefvooghe, B., & Vandierendonck, A. (2006). Stimulus- and response-conflict-induced cognitive control in the flanker task. *Psychonomic Bulletin & Review, 13*, 328–333. doi:10.3758/BF03193852
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*, 779–804. doi:10.3758/BF03194105
- Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*, 641–671. doi:10.1080/09541440802205067
- Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology, 46*, 361–413.
- Wendt, M., Kluwe, R. H., & Peters, A. (2006). Sequential modulations of interference evoked by processing task-irrelevant stimulus features. *Journal of Experimental Psychology. Human Perception and Performance, 32*, 644–667. doi:10.1037/0096-1523.32.3.644
- Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.
- Yantis, S., Meyer, D. E., & Smith, J. K. (1991). Analyses of multinomial mixture distributions: New tests for stochastic models of cognition and action. *Psychological Bulletin, 110*, 350–374. doi:10.1037/0033-2909.110.2.350