Psychonomic Bulletin & Review
2008, 15 (6), 1218-1228
doi:10.3758/PBR.15.6.1218

The EZ diffusion method: Too EZ?

Roger Ratcliff
Ohio State University, Columbus, Ohio

The diffusion model (Ratcliff, 1978) for fast two-choice decisions has been successful in a number of domains. Wagenmakers, van der Maas, and Grasman (2007) proposed a new method for fitting the model to data ("EZ") that is simpler than the standard chi-square method (Ratcliff & Tuerlinckx, 2002). For an experimental condition, EZ can estimate parameter values for the main components of processing using only correct response times (RTs), their variance, and accuracy, not error RTs or the shapes of RT distributions. Wagenmakers et al. suggested that EZ produces accurate parameter estimates in cases in which the chi-square method would fail—specifically, experimental conditions with small numbers of observations or with accuracy near ceiling. In this article, I counter these claims and discuss EZ's limitations. Unlike the chi-square method, EZ is extremely sensitive to outlier RTs and is usually less efficient in recovering parameter values, and it can lead to errors in interpretation when the data do not meet its assumptions, when the number of observations in an experimental condition is small, or when accuracy in an experimental condition is high. The conclusion is that EZ can be useful in the exploration of parameter spaces, but it should not be used for meaningful estimates of parameter values or for assessing whether or not a model fits data.

The diffusion model, as it is currently used in research with response time (RT) and accuracy as dependent measures, was first applied to two-choice decision data in cognitive tasks by Ratcliff (1978). The model explains RT and accuracy data jointly in terms of components of processing: the quality of the information on which a decision is based, the criterial amounts of information needed before a decision can be made, and the time taken by processes outside the decision itself. The model has been used to explain performance in a variety of two-choice tasks, and in each case it has accounted for the RT distributions of correct and error responses and the accuracy for each condition in an experiment.

The diffusion model assumes that evidence is noisy and is accumulated over time, from a starting point (z) toward one of two decision criteria. For example, if a subject is presented with a letter string and is asked to judge whether it is a word or a nonword, evidence is accumulated toward a "word" or a "nonword" boundary—a or 0, respectively. Evidence is represented by drift rate v, which in this example is positive for word stimuli and negative for nonword stimuli. Components of processing outside the decision process—such as encoding, lexical access, memory retrieval, response output, and so on—are combined into one parameter, T_{er}. The components of processing—v, z, and T_{er}—are assumed to vary from trial to trial, which adds three parameters to the model.

The diffusion model can be viewed as an extension/replacement for signal detection theory (SDT). In SDT, for each type of stimulus in an experiment, all sources of noise (e.g., perception, memory, criterion, and decision) are combined into a single value. In the diffusion model, four sources of noise are identified separately: variability within the decision process, as well as variability across trials in criteria, drift rate, and the nondecision component of processing. The across-trial variability in drift rate corresponds to noise in perception or memory, which is the source of noise usually assumed in SDT.

The standard way of fitting the diffusion model to data is to search for parameter values for the components of processing that best predict correct and error RT distributions and accuracy for all of the conditions in an experiment simultaneously. Maximum likelihood would be a more efficient method, but it is highly sensitive to contaminant RTs—that is, RTs that are not part of the decision process under investigation, such as for lapses of attention (Ratcliff & Tuerlinckx, 2002). This problem is so acute that a few bad data points can lead to estimates of parameter values an order of magnitude different from the true values. Ratcliff and Tuerlinckx showed that a chi-square method based on quantile RTs, which was subsequently used in research by Ratcliff and colleagues, does not have this problem; it is robust to small and even modest numbers of contaminant RTs (see the Appendix).

Wagenmakers, van der Maas, and Grasman (2007; hereafter, simply Wagenmakers et al.) proposed a simplified alternative method for estimating diffusion model parameters from data, the "EZ" method. According to this method, three statistics from a single experimental condition (accuracy and the mean and variance of the RTs for correct responses) are used to estimate v, a, and T_{er} for the condition. EZ assumes that the starting point is always halfway between the boundaries ($z = a/2$) and, unlike the full diffusion model (and SDT), that no across-trial variability in parameter values occurs. The assumption that drift rate is constant across trials means that if boundary separation is increased, accuracy for any condition will approach one. Wagenmakers et al. evaluated the EZ method fairly, with appropriate disclaimers about its applicability. In this article, I further evaluate EZ and show that it can produce misleading estimates of parameters.

There are several problems in applying EZ, as it has been implemented by Wagenmakers et al., to data. First, there is no foolproof way to test whether the EZ simplifications of the full diffusion model are valid. Second, as Wagenmakers et al. pointed out, the method can only fit one condition of an experiment at a time, so EZ cannot

R. Ratcliff, ratcliff.22@osu.edu

take advantage of across-condition constraints on data. Third, EZ uses only accuracy and the mean and variance of RTs for correct responses; it does not use error RTs or the shapes of the correct and error RT distributions. The shapes of RT distributions provide especially strong constraints on whether and how the diffusion model can account for empirical data (Ratcliff & Murdock, 1976; Ratcliff, Van Zandt, & McKoon, 1999; Ratcliff & McKoon, 2008). However, because distribution shape is relatively invariant across a large range of parameter values (see the discussion in Ratcliff & McKoon, 2008), the model will produce mismatches only in limited situations (see below). As a result, there is usually no way for EZ to detect mismatches between the diffusion model predictions and data. Thus, whether the diffusion model appropriately describes the data or is radically inconsistent with them, EZ will produce estimates of v, a, and T_{er}, so it will not be falsifiable without further evaluation, such as examining RT distributions. In contrast, the chi-square method requires the model to fit all aspects of the data and produces a goodness-of-fit value.

The chi-square and other quantile-based methods (e.g., G-square [Ratcliff & Smith, 2004] and QML [Heathcote, Brown, & Mewhort, 2002]) have become standard for applications of the diffusion model (e.g., Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Smith, 2004; Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2001, 2003, 2004, 2007; Thapar, Ratcliff, & McKoon, 2003). The latter studies found that small differences in parameter values between groups of subjects can be significant. Between 65- to 90-year-old subjects and college students, differences as small as 0.05 (15%) in v (drift rate), 40 msec in T_{er} (nondecision component), and 0.02 (15%) in a (boundary separation) have been significant. If EZ produces parameter estimates that consistently deviate from true values by much more than these differences, it can then lead to misinterpretations. The simulations of data described below show inaccuracy in EZ parameter estimates large enough to be of practical significance, large enough in fact that differences in performance between groups of subjects can be attributed to the wrong component of processing.

The goal of this article is to show when and on what bases misinterpretations can occur. In it, simulated data are used to test EZ in two ways. First, the effects of misspecifications and contaminant RTs on EZ estimates of parameter values are assessed, and second, the EZ and chi-square methods are compared in terms of the accuracy and efficiency with which parameter estimates are recovered from data. In other words, this article reports on how closely EZ estimates approach the true parameter values and how much variability occurs in the estimates.

In addition to evaluation of EZ, Wagenmakers et al.'s criticisms of the chi-square method are addressed. Specifically, contrary to Wagenmakers et al., the chi-square method is shown to give accurate parameter estimates for experimental conditions with highly accurate responses or small numbers of observations.

Contaminant Responses

Contaminant responses are a long-standing and difficult problem for experiments that use RT measures (see, e.g., Ratcliff, 1979, 1993; Ulrich & Miller, 1994). Contaminants are responses that do not derive directly from the processes under study; they include lapses of attention, spurious responses, gaps in processing, fast guesses that are not based on the stimulus (see, e.g., Swensson, 1972), and so on. Fast guess RTs are short, and most can be excluded from data analyses by removing responses faster than a cutoff value below which accuracy is at chance. Other contaminant responses are more difficult to deal with, however, because they are spread through the whole RT distribution; many occur in the tail of the distribution, but not all.

The effects of contaminant RTs on estimates of parameter values can be reduced by fitting the model to quantile RTs, as is done when the model is fit to data by the chi-square method. Using quantiles reduces the influence of a few outlier RTs because the quantiles are insensitive to RTs above or below them (Ratcliff & Tuerlinckx, 2002).

When contaminant RTs constitute more than a few percent of the data, they can nonetheless distort the fit of the model. To handle this problem, contaminant RTs can be explicitly modeled. Ratcliff and Tuerlinckx (2002) assumed that data contain contaminant responses and that these responses take extra time. These contaminants were modeled by assuming that they occur with probability p_o and increase RTs but do not affect accuracy (Table 7 of Ratcliff & Tuerlinckx shows recovered parameter values without explicitly modeling contaminants). For each condition in an experiment, the increase in RTs was assumed to be distributed uniformly through the RT distribution, with a range from the minimum to the maximum RT in a condition. To demonstrate that recovered parameter values were accurate for data with contaminants, Ratcliff and Tuerlinckx conducted simulations. Contaminants were simulated for 5% of the RTs by adding an amount of time between 0 and 2,000 msec to RTs generated from the model. This percentage is a reasonable upper limit, beyond which subjects would likely be eliminated from data analyses. With contaminants added in this way, the chi-square method accurately recovered the parameter values used to produce the data. Note that the assumption for generating contaminants in the simulated data was *not* the same as the assumptions used for fitting the model to data that include contaminants. The conclusion from the simulations was that, with the chi-square method, contaminants are not a problem for fitting data.

In contrast, contaminant RTs do present a problem for EZ, one that was not discussed by Wagenmakers et al. One of the inputs to EZ is the RT variance for correct responses. Variance estimates are extremely sensitive to contaminants and are poorly estimated from data (Ratcliff, 1979). For example, for an ex-Gaussian distribution of RTs with a mean of 700 msec and the shortest responses around 550 msec (a shape and values typical of recognition memory experiments; Ratcliff & Murdock, 1976), the standard deviation in the estimate of variance with 100 observations is about 25% of the total variance, and with 1,000 observations, it is about 10% (Ratcliff, 1979). This means that a 95% confidence interval around a variance estimate based on 100 observations would range from about one half to about one and a half times the true value. Because of this high sensitivity, the effects of contaminants on EZ can be large. Standard ways of eliminating slow outlier RTs from data can lead to radically different estimates of variance (e.g., Ratcliff, 1993). For example, in a recognition memory experiment with 13,600 observations, Ratcliff (1979) found that moving from an upper cutoff of 5 sec to one of 2 sec reduced the estimate of variance by almost a half, from 69,800 to 37,300 $msec^2$. This large difference resulted from elimination, between 2 and 5 sec, of only 0.8% of the data.

To demonstrate that contaminants have strong effects on EZ but not on the chi-square method, the two were applied to simulated data with and without contaminants (Table 1). The simulated data were generated from the model for three experimental conditions, with drift rates of 0.1 (a difficult condition), 0.3 (an easy condition), and 0.2. For each condition, the number of simulated trials was 10,000, a number that approximates exact predictions from the model. Contaminant RTs were generated by adding a uniformly distributed amount of time, with a minimum of 0 and a maximum of 2,000 msec, to RTs generated from the model (Ratcliff & Tuerlinckx, 2002); the probability of selecting a trial as an outlier was p_o. For generality in examining the effects of contaminants, another set of simulations was performed with the amount of added time generated from an exponential distribution with mean 1,000 msec.

The comparison of the chi-square method and EZ was biased in EZ's favor in three ways. First, following the EZ assumptions, the data were simulated with no across-trial variability in the values of v, T_{er}, or z. Second, z was set halfway between 0 and a ($z = a/2$). Third, the data for each experimental condition were fit separately, in accord with the inability of EZ to fit more than one condition at a time. To match EZ, the chi-square method was also fit to each experimental condition separately (although in practice, all of the conditions of an experiment would be fit simultaneously). Table 1 shows the results of the simulations when no data were excluded and when long RTs were excluded by an RT cutoff.

Without contaminants, EZ and the chi-square method both accurately recovered parameter values at the longer cutoffs, but as the cutoff was reduced, the EZ estimates began to move away from the true values to a greater degree than did the chi-square estimates. With 5% contaminants, the picture was different. With no cutoff, the EZ misestimates were large. As cutoffs were reduced, the estimates crossed the true values, but importantly, the cutoffs at which the parameter estimates matched the true values were different for the conditions with drift rate 0.3 and drift rate 0.1. The cutoffs for the best-matching estimates differed between the drift rates not only in terms of the RT above which data were eliminated, but also in terms of the proportion of contaminants eliminated and the number of standard deviations that the cutoff was above the mean RT. In contrast, the chi-square method estimated the true parameter values consistently well until the proportion of data eliminated began to exceed the proportion of contaminants (5%).

Even with few contaminants, the shapes of RT distributions can be badly mispredicted. Figure 1 shows a simulated RT distribution generated with 10,000 observa-

Table 1
Estimated Parameter Values From the EZ and Chi-Square Methods, With and Without Contaminants

	Contaminant	EZ			Chi-Square				Cutoff	Prop. of
v	Type and p_o	a	T_{er}	v	a	T_{er}	v	p_o	(msec)	Data Elim.
0.1	Rectangular,	0.118	0.301	0.098	0.118	0.321	0.102	.001	5,000	.0000
	$p_o = 0$	0.117	0.306	0.099	0.118	0.321	0.102	.000	2,000	.0008
		0.111	0.326	0.104	0.116	0.324	0.104	.000	1,500	.0113
		0.094	0.360	0.122	0.104	0.332	0.114	.000	1,000	.0784
0.1	Rectangular,[1]	0.142	0.216	0.082	0.119	0.317	0.101	.042	5,000	.0000
	$p_o = .05$	0.125	0.286	0.093	0.119	0.317	0.102	.037	2,000	.0174
		0.114	0.322	0.101	0.119	0.317	0.101	.000	1,500	.0379
		0.095	0.361	0.121	0.105	0.333	0.114	.000	1,000	.1142
0.3	Rectangular,	0.119	0.301	0.298	0.117	0.312	0.305	.000	5,000	.0000
	$p_o = 0$	0.117	0.303	0.303	0.118	0.299	0.304	.011	1,500	.0001
		0.111	0.314	0.307	0.116	0.301	0.302	.000	1,000	.0055
		0.105	0.326	0.339	0.113	0.330	0.336	.000	800	.0311
0.3	Rectangular,[1]	0.174	0.125	0.196	0.117	0.305	0.304	.057	5,000	.0000
	$p_o = .05$	0.150	0.209	0.226	0.116	0.306	0.302	.048	2,000	.0122
		0.129	0.273	0.264	0.117	0.300	0.301	.032	1,500	.0443
		0.113	0.311	0.301	0.116	0.300	0.304	.015	1,000	.0738
0.2	Exponential,	0.178	0.052	0.134	0.118	0.305	0.197	.043	5,000	.0055
	$p_o = .05$	0.129	0.270	0.181	0.117	0.314	0.200	.033	2,000	.0124
		0.121	0.297	0.194	0.117	0.318	0.202	.029	1,500	.0320
		0.105	0.333	0.222	0.113	0.337	0.213	.000	1,000	.0710

Note—Each condition was fit separately for both methods. The parameter values used to generate the simulated data were $a = 0.12$, $T_{er} = 0.3$, and $z = a/2$. a, boundary separation; T_{er}, nondecision component of response time; v, drift rate; p_o, proportion of contaminants. [1]For these conditions with contaminants, the 2*SD* and 3*SD* cutoffs were 1,380 and 1,738 msec, for $v = 0.1$, and 1,142 and 1,443 msec, for $v = 0.3$.

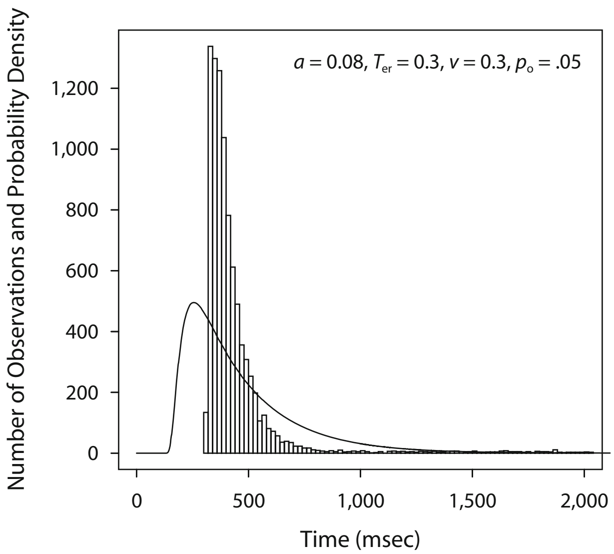

Figure 1. RT distribution histogram from the diffusion model, with 5% outliers from a uniform distribution with range 2,000 msec. The parameter values are shown in the figure; in addition, $z = a/2$, and the variability across-trial parameters are set to zero. The continuous distribution is the distribution from the parameter values from the EZ method fit.

tions and 5% contaminants, along with the RT distribution predicted from the EZ parameter values. Because slow contaminants produce a large increase in variance relative to the mean, the EZ-predicted distribution is considerably wider than the distribution for the simulated data. The misses between data generated from EZ parameters (e.g., those in Figure 1) and real data can be large in chi-square terms. For example, one of the larger misses by EZ (row 5 of Table 1), a case with 5% contaminant RTs, produced predicted quantiles that missed the real data by about 50 msec for the .1 quantile and about 40 msec for the .9 quantile. The resulting chi-square value was 1,007, 66 times larger than when the chi-square method's parameter values were used to generate predicted data.

Results were similar when contaminant RTs were generated from an exponential instead of a uniform distribution (bottom rows in Table 1). The parameter values recovered with the chi-square method were approximately accurate, whereas those recovered with EZ significantly missed the true values with cutoffs of 5,000, 2,000, and 1,000 msec. The reason that the chi-square fitting method works for exponentially distributed contaminants is that the addition of contaminants to the diffusion model's RT distribution (a convolution) adds to the tail of the distribution, with little difference from what is added under a uniform distribution. The exponential assumption places more of the distribution of contaminants earlier in the regular RT distribution, relative to the uniform assumption, so the proportion of contaminants is slightly underestimated.

Summary. Real experiments contain contaminant RTs, so it is worth stressing that the problems presented by contaminants for the EZ method can be serious. For a relatively difficult condition (the 0.1 drift rate above), cutoffs of 5,000 and 1,000 msec yielded misses in parameter estimates ranging from 12% to 20%. For an easier condition (the 0.3 drift rate above), the misses with 1,500-, 2,000-, and 5,000-msec cutoffs ranged from 45% to 58%.

The results of the simulations also show that no simple rule will improve EZ performance when the data contain contaminants and that the EZ parameter estimates are a function of the cutoff used. For the 0.1 drift rate condition, optimal EZ estimates were obtained with a cutoff somewhere between 2,000 msec and 1,500 msec (about 3 *SD*s above the mean, eliminating about 2.5% of the data). But for the 0.3 drift rate condition, optimal estimates were obtained with a cutoff near 1,000 msec (about 2 *SD*s above the mean, eliminating about 7% of the data). Simple rules such as cutting out responses longer than *X* msec, cutting out the slowest *Y*% of responses, or cutting out responses longer than *Z SD*s above the mean will fail to yield accurate parameter estimates across all of the conditions in an experiment. Inaccurate estimates can, in turn, lead to misinterpretations of the data. The chi-square method, in contrast, by modeling contaminants, can accommodate data that include contaminants, and thus can reduce or eliminate their effects on parameter estimates.

Experimental Conditions With High Accuracy

Wagenmakers et al. made a particularly strong claim about experimental conditions with high accuracy. They claimed that, for such conditions, too few errors are present to provide the error RT distributions required by the chi-square method, and that the number of errors required to accurately estimate error RT distributions is larger than would be feasible for most experiments. For this reason, they said that accurately estimating parameters with the chi-square method might not be possible in most practical settings (p. 7). However, this claim is contradicted by published fits of the model to over 500 individual subjects' data (e.g., Ratcliff, Thapar, et al., 2004; Ratcliff, Thapar, & McKoon, 2001, 2003, 2004, 2007; Thapar et al., 2003).

The reason that high-accuracy conditions are not problematic for the chi-square method is that all of the conditions of an experiment are fit simultaneously, therefore taking advantage of the mutual constraints on parameter values from one condition to another. In contrast, EZ currently can only fit the data from each condition of an experiment separately (e.g., in a lexical decision study, fit to word and nonword responses separately). Wagenmakers et al. discuss possibly constraining parameters across conditions by setting them equal and then fitting the data by weighted least squares. But this requires the kind of model-fitting procedure EZ is designed to avoid.

Table 2 shows the strength of the constraints on parameter estimates from one condition to another. Exact predictions from the model were generated using numerical solutions in order to give accuracy and RT quantiles for four experimental conditions. Parameters were set at values similar to those from previous empirical applications of the model, and the data included 5% contaminants (chosen from a uniform distribution). For three conditions, data were generated with drift rates of 0.1, 0.3, and −0.3. The probabilities of correct responses were .73, .96, and .96, respectively. For the fourth condition, the drift rate was 0.6 and the probability of a cor-

Table 2
True Parameter Values and Parameter Values Estimated by the Chi-Square Method

	a	T_{er}	η	s_z	v_1	v_2	v_3	v_4	z	p_o	s_t
True: One extreme drift	0.12	0.30	0.08	0.02	0.30	0.10	−0.30	0.60	0.06	.05	0.10
Recovered: One extreme drift	0.120	0.30	0.080	0.020	0.300	0.100	−0.300	0.601	0.060	.049	0.101

Note—a, boundary separation; T_{er}, nondecision component of response time; v, drift rate; η, standard deviation in drift across trials; s_z, range of distribution of starting point (mean $z = a/2$); p_o, proportion of contaminants; s_t, range of distribution of nondecision times.

rect response .9986 (i.e., 1 error in 1,000 observations, on average). The chi-square method succeeded in accurately estimating the parameter values, within 2% of their true values, as shown in Table 2. The conditions with drift rates of 0.1, 0.3, and −0.3 constrained the estimates of a, z, T_{er}, and the across-trial variability parameters. Because of these constraints, RT quantiles alone were sufficient to recover the drift rate for the high-accuracy condition. The power that this gives the model has been demonstrated in many not-yet-published experiments, including Ferrera, Grinband, Xiao, Hirsch, and Ratcliff (2006) and White, Ratcliff, Vasey, and McKoon (in press). For example, in the Ferrera et al. experiments, as many as half of the experimental conditions had no errors (out of a total of 24 or 30 conditions).

These simulations demonstrate that, contrary to Wagenmakers et al.'s claim, the chi-square method can be used to estimate parameters for experimental conditions with high accuracy. The only requirement is that the experiment must include conditions with lower accuracy values. In lexical decision, for example, accuracy is typically lower for very-low-frequency words. In perceptual experiments, it is usually possible to include conditions for which accuracy is near chance.

In contrast, EZ cannot recover parameter values when accuracy is at ceiling. Wagenmakers et al. (p. 11) suggested applying a standard correction to address this problem, replacing 1.0 accuracy values by 1.0 minus half of the minimum step below 1.0—that is, $1 - (0.5/N)$, where N is the number of observations. However, this can produce incorrect parameter values if true accuracy is considerably higher than $1 - (0.5/N)$. This is shown for the experimental condition in Table 2 with $a = 0.12$, $T_{er} = 0.3$, and $v = 0.6$, with predicted accuracy of .9986. With $N = 50$, accuracy would be estimated at 1.0, and application of the correction would give .99. When EZ was then applied with accuracy at .99 for the simulated data, the estimated parameter values missed the true values: $a = 0.088$ (a serious miss), $T_{er} = 0.317$, and $v = 0.523$.

The Accuracy and Efficiency of Parameter Estimates

In the section on contaminant RTs above, EZ produced inaccurate parameter estimates when the data contained contaminants. In this section, efficiency and accuracy will be considered. Data were simulated for four conditions, with drift rates of 0.3, 0.1, −0.1, and −3.0. There were three sets of simulations, and in each case a simulation was repeated 40 times in order to determine the *SD*s in the parameter estimates that would be expected across 40 experiments. The conditions were fit simultaneously for the chi-square method and separately for the EZ method, but a and T_{er} were averaged over all conditions (i.e., 4 × 40 values) for the EZ method.

No across-trial variability. Data were generated with $z = a/2$, no contaminants, and no across-trial variability in any of the parameters. The chi-square and EZ methods estimated drift rates reasonably accurately and with about the same amount of variability. However, the estimates for a and T_{er} had 3–5 times less variability for the chi-square than for the EZ method. The constraints across conditions led to much better estimation of a and T_{er} for the chi-square method.

Part of the poor efficiency of the estimates of a and T_{er} in the EZ method is due to correlations in parameter values. The variance for RT distributions is rather poorly estimated (Ratcliff, 1979) relative to mean RT and accuracy. The values of a and T_{er} are determined to a large degree by the variance, and if by chance it is high, a is estimated to be higher and T_{er} to be lower than the values used to generate simulated data. The correlation between the variance and a is about .9, that between the variance and T_{er} is about −.95, and that between a and T_{er} is about −.9 for the typical values used in this article (e.g., those in Table 3). The correlations between parameters from the chi-square method are much smaller, and the correlations between chi-square and EZ parameters are small. This suggests that parameter estimates from the EZ and chi-square methods are sensitive to different aspects of the data.

Across-trial variability in parameters. The simulated data for this analysis were more realistic and were generated with contaminant RTs and $z = a/2$. The parameter values were typical of those obtained from the Ratcliff, Thapar, et al. experiments (e.g., Ratcliff, Thapar, et al., 2004; Ratcliff, Thapar, & McKoon, 2001, 2003, 2004, 2007; Thapar et al., 2003); the amounts of across-trial variability were at the low end of the range that has been obtained empirically, and the proportion of contaminants was at the upper end. Wagenmakers et al. pointed out the problems for EZ when drift rate varies across trials. The results of the simulations here demonstrate the size of the misses that EZ can produce, as compared with the chi-square method, when EZ is seriously misspecified.

The middle section of Table 3 shows the true parameter values along with the means and *SD*s across the 40 experiments for the recovered values. With the EZ method, the misses are sufficiently large that their *SD*s (i.e., efficiency) are largely irrelevant. In contrast, the chi-square method recovers parameters reasonably well, about as well as when the data have no contaminants and no across-trial variability (top section of Table 3). Drift rate, variability in drift rate across trials, and variability in starting point across trials were all overestimated, but the estimates were

Table 3
Efficiency: Means and Standard Deviations in Parameters Estimated by the EZ and Chi-Square Methods

Method		a	T_{er}	η	s_z	v_1	v_2	v_3	v_4	z	p_o	s_t
True parameters		0.12	0.30	0.00	0.00	0.30	0.10	−0.10	−0.30	0.06	–	0.00
EZ[1]	*M*	0.118	0.300			0.303	0.100	−0.102	−0.300			
	SD	0.006	0.015			0.015	0.010	0.009	0.019			
Chi-square[2]	*M*	0.117	0.299			0.293	0.100	−0.098	−0.288	0.059	.001	
	SD	0.002	0.003			0.016	0.009	0.010	0.013	0.002	.005	
Chi-square[3]	*M*	0.118	0.305	0.032	0.021	0.313	0.106	−0.105	−0.307	0.060	.004	0.042
	SD	0.004	0.006	0.034	0.016	0.020	0.013	0.013	0.025	0.003	.010	0.034
True parameters		0.12	0.30	0.08	0.02	0.30	0.10	−0.10	−0.30	0.06	.05	0.10
EZ[1]	*M*	0.171	0.116			0.185						
		0.137	0.211				0.072					
	SD	0.012	0.034			0.012						
		0.018	0.032				0.009					
Chi-square[3]	*M*	0.129	0.304	0.133	0.046	0.344	0.120	−0.120	−0.340	0.065	.043	0.093
	SD	0.010	0.012	0.050	0.030	0.059	0.024	0.024	0.056	0.005	.019	0.031
True parameters		0.12	0.30	0.00	0.00	0.15	0.20	−0.20		0.06	–	0.00
EZ[4]	*M*	0.120	0.307			0.163	0.202	−0.203				
	SD	0.018	0.054			0.051	0.012	0.011				
Chi-square[3]	*M*	0.120	0.310	0.055	0.032	0.161	0.219	−0.223		0.060	.006	0.045
	SD	0.006	0.010	0.036	0.025	0.068	0.027	0.024		0.003	.011	0.033

Note—There were 400 observation per condition, except for v_1 ($N = 25$) in the lower five rows of the table. Also, $z = a/2$. a, boundary separation; T_{er}, nondecision component of response time; v, drift rate; η, standard deviation in drift across trials; s_z, range of the distribution of starting point (z); p_o, proportion of contaminants; and s_t, range of the distribution of nondecision times. [1]The means and *SD*s in the parameters a and T_{er} for the EZ method were computed from all the values from all four separate conditions. [2]η, s_z, and s_t were fixed at zero. [3]η, s_z, and s_t were allowed to vary freely. [4]The *SD*s in a and T_{er} for the EZ method in the bottom section are for the $N = 25$ condition. The *SD*s in a and T_{er} for the $N = 400$ condition were about the same as those in row 3.

within 1 *SD* of their true values (see Ratcliff & Tuerlinckx, 2002, for explanations).

One way of determining whether across-trial variability parameters are needed is to look at whether the chi-square value decreases significantly with the addition of nonzero variability parameters (see, e.g., Vandekerckhove & Tuerlinckx, 2007). The average chi-square across the 40 experiments with zero variability was 60.8 with 36 degrees of freedom. The average chi-square with nonzero variabilities was significantly different, 38.1 with 33 degrees of freedom. This result also indicates that the method of applying the model with and without nonzero parameters can be used to hierarchically assess the need for specific parameters and for other assumptions (e.g., equal durations of T_{er} in a speed–accuracy manipulation).

Small numbers of observations. Wagenmakers et al. claimed that the chi-square method cannot produce accurate estimates of parameters when the number of observations in an experimental condition is small. In evaluating this claim, it is important to consider the kind of experimental design that would give rise to a small number of observations. First, no experiment would ever have only one condition (e.g., only words or only nonwords in a lexical decision experiment). Moreover, the diffusion model would never be applied to such an experiment, because it is a two-choice, not a one-choice, model (see Gomez, Ratcliff, & Perea, 2007). Second, it would be unusual to run an experiment with only, say, 50 trials, 25 with one response and 25 with the other. In fact, the first 50 trials of an experiment are often used as practice, often with contaminant RTs, as subjects become familiar with the experimental task. More typically, a condition with a small number of observations would be embedded in a larger design with larger numbers of observations in other conditions and also with filler items (see White et al., in press). For example, in a lexical decision experiment, 25 items in a condition of interest might be embedded in a list of 400 filler words and 400 filler nonwords. The total of 825 trials, each taking 1.5 sec, would require only about 30 min of testing.

To show that the chi-square method can be applied to conditions with small numbers of observations, data were simulated for two experimental conditions with 400 observations each and one condition with 25 observations. Both the EZ and chi-square methods were applied to the data, and the chi-square method was fit to all of the conditions simultaneously. The simulations were set to favor EZ: zero across-trial variability, no contaminants, and $z = a/2$.

The chi-square method produced reasonably accurate estimates of a and T_{er}, with much lower standard deviations across simulations than those from EZ, by factors of 3.0 and 5.4, respectively (bottom section of Table 3). Contrary to Wagenmakers et al.'s claim, the chi-square method also provided a reasonably accurate estimate of drift rate for the condition with only 25 observations, although it was slightly less efficient (1.3 times larger) than EZ. The chi-square method could be made more efficient for drift rate if the across-trial variability parameters were fixed at zero (rows 4 and 5 of Table 4). The chi-square method would do much better than EZ if there were contaminant RTs in the data or variability across trials in drift rates.

The simulations in Table 4 were conducted in order to assess what happens when an experiment has two conditions and both have low numbers of observations (50 in each condition). The simulations were run with and without across-trial variability in parameters, with $z = a/2$, and

Table 4
Efficiency: Means and Standard Deviations in Parameters Estimated by the EZ and Chi-Square Methods With Low N (2 Conditions With $N = 50$ per Condition)

Method		a	T_{er}	η	s_z	v_1	v_2	z	s_t
True parameters		0.12	0.30	0.00	0.00	0.20	−0.20	0.06	0.00
EZ[1]	M	0.123	0.297			0.212	−0.218		
	SD	0.021	0.040			0.039	0.042		
Chi-square[2]	M	0.118	0.296			0.184	−0.190	0.059	
	SD	0.010	0.014			0.042	0.045	0.005	
Chi-square[3]	M	0.130	0.320	0.088	0.041	0.257	−0.259	0.065	0.087
	SD	0.020	0.026	0.070	0.046	0.073	0.087	0.012	0.059
True parameters		0.12	0.30	0.16	0.02	0.20	−0.20	0.06	0.10
EZ[1]	M	0.110	0.283			0.151	−0.144		
	SD	0.013	0.036			0.035	0.031		
Chi-square[3]	M	0.132	0.311	0.230	0.045	0.242	−0.217	0.066	0.092
	SD	0.016	0.025	0.075	0.045	0.076	0.076	0.011	0.061

Note—$z = a/2$. a, boundary separation; T_{er}, nondecision component of response time; v, drift rate; η, standard deviation in drift across trials; s_z, range of the distribution of starting point (z); and s_t, range of the distribution of nondecision times. [1]The means and SDs in the parameters a and T_{er} for the EZ method were computed from all the values from the two separate conditions. [2]η, s_z, and s_t were fixed at zero. [3]η, s_z, and s_t were allowed to vary freely.

without contaminants. The patterns here are the same as in Table 3. When there was no across-trial variability and the variability parameters were set to zero, the chi-square method gave values of a and T_{er} that were more efficient than those from EZ. When the variability parameters were free to vary, the chi-square estimates were biased away from the true values and the variability parameters were estimated to be greater than zero. When the data were simulated with across-trial variability, EZ and the chi-square method produced similar magnitudes of biases in parameter values. In these two cases, though, the standard deviations in the values were large, and neither method provided adequate estimates of parameter values.

When the total number of observations is small in every condition of an experiment and there is across-trial variability, both EZ and the chi-square method will produce biased estimates. Also, if some of the subjects make no errors, neither method can then be trusted. For the chi-square method, the recommendation is to always test conditions with small numbers of observations in the context of conditions with large numbers. The large-number conditions will then provide the necessary constraints to obtain accurate parameter estimates for the condition with few observations.

Misspecification

Wagenmakers et al. suggested several tests for determining when EZ should not be applied to data—that is, when EZ's assumptions are not met. First, the RTs of correct and error responses for a condition should not be significantly different from each other. If they are different, then η and s_z are different from zero, counter to the rule for applying EZ. The second test compares RTs for the two responses, positive and negative (e.g., word and nonword in lexical decision). EZ should not be applied when corrects are faster than errors for one response but slower than errors for the other. This pattern of data would arise when $z \neq a/2$ (Wagenmakers et al., p. 17).

The problem with these tests, and a major problem for EZ, is that for conditions with high accuracy or with small numbers of observations—exactly the conditions for which EZ might have an advantage—there will be too little power to determine whether correct and error RTs are significantly different. Consider, for example, an experimental condition with $a = 0.12$, $T_{er} = 0.3$, $\eta = 0.16$, $s_z = 0.02$, $v = 0.3$, $s_t = 0.2$, and $z = a/2$ (all of which are consistent with fits to real data) and no contaminants. The mean RT for correct responses is 494 msec, the mean RT for errors is 595 msec, the proportion of correct responses is .915, and the proportion of errors is .085. The standard deviation in the correct RTs is 192 msec, and the standard deviation in the error RTs is 231 msec. For the difference in the two mean RTs to be significant, the difference would have to be at least twice the standard error in the error RTs. For this to be the case, the experimental condition would need a relatively large number of observations, at least 250.

Also, because variability in drift rate across trials leads to errors slower than correct responses, whereas variability in starting point leads to faster errors, combinations of these can produce little or no difference between correct and error RTs. Because only drift rate variability across trials leads to incorrect parameter estimates, such combinations can defeat the test for misspecification.

Another possible misspecification concerns EZ's assumption that the starting point is equidistant from the boundaries ($z = a/2$). Wagenmakers et al. noted that inaccurate estimates might occur when this assumption is violated, but they did not show how large these misses might be. To address this issue, simulated data were generated with z at either $a/4$ or $3a/4$ and with two drift rates, $v = 0.1$ and $v = 0.3$. Data with values of z this extreme can occur when the probability of one of the responses is much greater than the probability of the other (e.g., 80:20). To produce parameter estimates as exact as possible, 10,000 data points were generated for each condition. There were no contaminants in the data, nor was there across-trial variability in any of the parameters.

As shown in Table 5, EZ's estimates deviated from the true values. For $v = 0.1$, the estimates of a were close to the true values, but for $v = 0.3$, they deviated by about 0.02. The estimates of T_{er} deviated from the true values for both $v = 0.1$ and 0.3, by between 70 and 90 msec.

Table 5
Estimated Parameter Values With $z \neq a/2$

True Values		Parameter Values From EZ			Parameter Values From Chi-Square			
v	z	a	T_{er}	v	a	T_{er}	v	z
0.1	0.03	0.126	0.202	0.191	0.118	0.313	0.102	0.028
0.1	0.09	0.114	0.371	−0.007	0.118	0.310	0.097	0.088
0.3	0.03	0.140	0.227	0.414	0.123	0.308	0.301	0.026
0.3	0.09	0.092	0.379	0.166	0.118	0.300	0.303	0.089

Note—The parameter values used to generate the simulated data were $a = 0.12$ and $T_{er} = 0.3$.

The drift rate estimates differed from the true values by between 0.09 and 0.14. In contrast, the chi-square method gave reasonably accurate estimates of all parameters.

Misinterpretations of Experimental Effects

Perhaps the most serious problem with EZ is that for some patterns of empirical data, it leads to misinterpretations of the effects of independent variables. Here, three plausible examples are described, all framed in terms of differences in cognitive processing between older adults and college students, using parameter values chosen to be typical of those from experiments by Ratcliff and colleagues (e.g., Ratcliff, Thapar, et al., 2004; Ratcliff, Thapar, & McKoon, 2001, 2003, 2004, 2007; Thapar et al., 2003). The important point of the simulations is that EZ does not correctly identify the sources of the differences between the two groups. For each of the simulations, 10,000 data points were generated, with the parameter values shown in Table 6.

For the first example, the simulated data were generated to produce worse performance for the older adults by setting their boundary separation wider, their nondecision components slower, and their drift rate lower (see, e.g., Thapar et al., 2003). College students' performance was better in every way, except that they had occasional lapses of attention, such that 1% of their responses were contaminants (from a uniform distribution). This difference is plausible, because older subjects "try harder" than college students and therefore are less likely to have contaminants such as lapses of attention. There was no variability in parameter values across trials, and the starting point was equidistant between the two boundaries. EZ parameter estimates gave an incorrect picture of the differences between the two groups. According to the EZ estimates, the only difference between the groups was that T_{er} was longer for the older adults. Wagenmakers et al. suggested that EZ mistakes like this could be avoided by using their tests for misspecification, but data in this example would pass the tests.

In the second example, college students and older adults did not differ in a, T_{er}, or v. The only difference between the two groups was that the older adults' components of processing were moderately variable across trials, but the college students' were not. There were no contaminant RTs, and $z = a/2$. EZ incorrectly estimated larger drift rates for the college subjects. This mistaken application could not be determined from the relative speeds of correct versus error responses; the RTs were equal at 502 msec. This occurs because variability in drift and in starting point across trials leads to errors that are slower than correct responses and faster than correct responses, respectively, so that the errors balance each other out (see, e.g., Experiment 1 of Ratcliff & McKoon, 2008).

In the third example, the older adults had the same boundary separation and the same drift rate as the college students. There was no across-trial variability for either group, and there were no contaminant RTs. The two groups differed only in starting point; the older subjects were slightly biased toward the top boundary (0.045 instead of the unbiased value, 0.04), and the college students were unbiased. This bias is low relative to manipulations of response proportions in real data (see, e.g., Ratcliff & Smith, 2004, Experiment 3). Again, EZ produced a misinterpretation of the data, attributing the difference in performance between the groups to a lower drift rate for the older subjects instead of to a difference in starting point. The data indicated only a minor misspecification. Error RTs were only 22 msec longer than correct RTs (499 vs. 521 msec), and this difference would almost certainly not be large enough for a researcher to rule out application of the EZ method.

These examples show that RT differences said to signal when EZ should not be applied can be insignificant, even when they are large enough that EZ produces misinterpretations. For applications to real data, to determine how bad a misinterpretation might be, simulations should be conducted with appropriate parameter values, numbers of observations, and standard deviations in correct and error RTs for each of the two responses in each condition of an experiment.

Applications to real data. Wagenmakers et al. compared the EZ and chi-square methods for one set of data from a two-choice perception task. They found that the estimated values from EZ differed by as much as 25% from the chi-square estimates (their Figure 12).

Table 6
Examples in Which EZ Misspecification Leads to Incorrect Interpretations

		True Parameters							EZ Estimates		
Simulation	Subjects	a	T_{er}	v	z	η	s_z	p_o	a	T_{er}	v
1: Contaminants, $p_o = .01$	Older	**0.135**	**0.44**	**0.25**	**0.0675**	0.00	0.00	**.00**	0.132	0.445	0.249
	College	**0.120**	**0.40**	**0.30**	**0.0600**	0.00	0.00	**.01**	0.138	0.337	0.247
2: Drift & starting pt. variability	Older	0.080	0.40	0.30	0.0400	**0.12**	**0.04**	.00	0.076	0.388	0.240
	College	0.080	0.40	0.30	0.0400	**0.00**	**0.00**	.00	0.078	0.402	0.305
3: Starting pt. bias	Older	0.080	0.40	0.30	**0.0450**	0.00	0.00	.00	0.074	0.411	0.259
	College	0.080	0.40	0.30	**0.0400**	0.00	0.00	.00	0.078	0.401	0.305

Note—Differences in parameter values between simulated older subjects and college students are in bold.

Table 7
EZ Fits for Brightness Discrimination (Ratcliff et al., 2003)

Model	Subjects	Prop. of Data Elim.	Speed Instructions			Accuracy Instructions		
			a	T_{er}	v	a	T_{er}	v
Standard	Young		0.073	0.405	0.264	0.135	0.405	0.264
	Older		0.074	0.459	0.272	0.122	0.459	0.272
EZ	Young	.0002	0.085	0.341	0.160	0.143	0.294	0.148
		.008	0.082	0.357	0.163	0.130	0.366	0.158
		.022	0.079	0.369	0.167	0.114	0.419	0.178
	Older	.0003	0.088	0.380	0.196	0.113	0.415	0.195
		.004	0.083	0.405	0.201	0.108	0.444	0.199
		.012	0.081	0.414	0.205	0.101	0.469	0.209

To provide another real-data comparison, the two methods were applied to a brightness discrimination experiment with 60- to 75-year-olds and college students (Ratcliff et al., 2003; see also Ratcliff, 2002). The difficulty of the task was manipulated with levels of brightness and presentation duration, and subjects were tested with both speed and accuracy instructions.

Fitting the data with the chi-square method (Table 7), the surprising result was a finding of no significant differences in drift rates between the young and older subjects (replicated in Ratcliff, Thapar, & McKoon, 2006). EZ estimates led to a different conclusion. The EZ parameters were computed for each experimental condition for each subject and then averaged. At three RT cutoffs, drift rates differed significantly between the young and older subjects ($p < .001$ for paired t tests on the average drift rates across subjects, for the 18 conditions formed by crossing brightness, duration, and speed vs. accuracy instructions).

For some experiments, the parameter values estimated by the two methods might be close enough to lead to the same conclusions (e.g., boundaries wider for one group relative to another). But for the example just described, EZ and the chi-square method produced divergent interpretations of the data. Thus, it is difficult to argue that EZ is an "attractive alternative" (Wagenmakers et al., p. 20) to the chi-square method. It is important to reiterate that, if the across-trial variability parameters had been zero, the chi-square method would have produced parameter values close to the EZ values, because EZ is nested in the full model. It is also important to note that EZ tests for misspecification would have produced a warning that EZ was inappropriate (error RTs were slower than correct RTs—overall, 30 msec slower—for some proportion of subjects, as in Wagenmakers et al.'s experiment). However, it would be easy for a researcher to miss the warning or ignore it, because the high correlation between EZ and full-model parameters in Wagenmakers et al.'s experiment seems to suggest that the two methods will produce the same interpretation of the data.

Limitations of the Chi-Square Method and the Diffusion Model

Wagenmakers et al. criticized the diffusion model in several ways. Most of the criticisms were not sufficiently evaluated, and several were invalid. First, Wagenmakers et al. (p. 7) claimed that "when a model with at least seven free parameters is unleashed on a small data set, problems such as high-variance parameter estimates and sensitivity to starting points may become prominent." In contradiction, the simulations in Tables 1–5 show that in realistic situations, in which there are contaminant RTs and across-trial variability in processing, the chi-square method is accurate in recovering parameter values and reasonably efficient. Also, in practice, as long as T_{er} is placed appropriately (e.g., about 50 msec less than the average .1 quantile RT in the data) and drift rates have the correct sign (positive or negative), recovered parameter values are almost always close to the best values.

Wagenmakers et al. (p. 19) also implied that the diffusion model is prone to overfitting, writing that simpler models "are less prone to overfitting (i.e., modeling noise)." In other words, if a model is too complex, the parameters estimated from data can follow noise in the data and be biased away from the parameters' true values. It turns out that the full diffusion model will overfit somewhat, but only with lower numbers of observations, whereas the EZ method will overfit to accommodate random variation in the estimate of RT variance.

Wagenmakers et al. argued that the diffusion model is limited in that it does not explain how subjects set or adjust decision criteria. This criticism is valid, but it applies to any model with decision criteria, including SDT. There currently appear to be no satisfactory general models for criterion setting. Such models are hard to develop because human subjects, given verbal instructions, can often accurately calibrate their criteria even for the first trial of an experiment.

Wagenmakers et al. also complained that the diffusion model includes no model of sequential effects and does not explain why the stimuli of one condition are more difficult than the stimuli of another. But they miss the central point that the diffusion model provides a meeting ground between models. For example, the diffusion model can estimate parameter values conditionalized on prior stimuli, and so provide parameter values to use in models of sequential effects. Likewise, a model for stimulus encoding must provide a drift rate such that when that drift rate drives the diffusion decision process, the correct predictions for RT distributions and accuracy are obtained (some models do exactly this: Ratcliff, 1981; Smith, Ratcliff, & Wolfgang, 2004).

Discussion and Conclusions

The EZ method can be useful for exploring the behavior of the diffusion model and how it might account for various experimental effects, and can also provide a first pass at experimental data. However, the simulations presented in this article and the analysis of the data from Ratcliff et al. (2003) show that EZ and the chi-square method can lead to different estimates of parameter values, and so to different interpretations of experimental effects.

For the simulated data, EZ consistently produced less accurate parameter estimates than did the chi-square method, with a wider margin of error. In practical settings, contaminant RTs will almost certainly lead EZ to incorrect parameter estimates for some or all of the conditions in an experiment, and the standard methods of eliminating contaminants will not solve the problem. In consequence, EZ interpretations of data can be wrong, attributing the effects of an independent variable to the wrong component of processing.

The chi-square method can produce accurate estimates of parameter values even when the number of observations in a condition is small or when a condition has high accuracy (counter to Wagenmakers et al.'s claim). The only requirement is that an experiment include other conditions (or filler items) with several hundred observations and that these conditions provide sufficient errors to estimate RT distributions for errors. These other conditions constrain the parameter values so that quantile RTs alone are sufficient to estimate drift rate for either a high-accuracy or small-N condition. Currently, EZ is fit to the data for each condition of an experiment separately, so it cannot take advantage of across-condition constraints.

It is important to emphasize that EZ parameter estimates can be wrong, even when the data pass Wagenmakers et al.'s tests for misapplication of EZ. In the circumstances in which EZ would be most useful—small numbers of observations and high accuracy—the power of the tests will be extremely low. Most data sets will pass the tests, whether or not EZ is appropriate. The conclusion is that there is no substitute for collecting sufficient data and modeling the data with a method that provides accurate estimates of parameters.

AUTHOR NOTE

Preparation of this article was supported by NIMH Grant R37-MH44640, NIA Grant R01-AG17083, and NIDCD Grant R01-DC01240. I thank Andrew Heathcote and Gail McKoon for extensive comments on this article. Correspondence relating to this article may be sent to R. Ratcliff, Department of Psychology, Ohio State University, 1835 Neil Ave., Columbus, OH 43210 (e-mail: ratcliff.22@osu.edu).

REFERENCES

Ferrera, V. P., Grinband, J., Xiao, Q., Hirsch, J., & Ratcliff, R. (2006, October). *Distinguishing evidence accumulation from response bias in categorical decision making*. Poster presented at the annual meeting of the Society for Neuroscience, Atlanta, GA. Abstract available at www.sfn.org.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, **136**, 389-413.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, **9**, 394-401.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308-313.

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

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446-461.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, **88**, 552-572.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, **114**, 510-532.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, **9**, 278-291.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, **111**, 159-182.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, **20**, 873-922.

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**, 190-214.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, **111**, 333-367.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology & Aging*, **19**, 278-289.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology & Aging*, **16**, 323-341.

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics*, **65**, 523-535.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory & Language*, **50**, 408-424.

Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review*, **13**, 626-635.

Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychology & Aging*, **22**, 56-66.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, **9**, 438-481.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, **106**, 261-300.

Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research*, **44**, 1297-1320.

Swensson, R. G. (1972). The elusive tradeoff: Speed vs. accuracy in visual discrimination tasks. *Perception & Psychophysics*, **12**, 16-32.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology & Aging*, **18**, 415-429.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, **123**, 34-80.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, **14**, 1011-1026.

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, **14**, 3-22.

White, C., Ratcliff, R., Vasey, M., & McKoon, G. (in press). Dysphoria and memory for emotional material: A diffusion model analysis. *Cognition & Emotion*.

(*Continued on next page*)

APPENDIX
The Chi-Square Method for Fitting the Diffusion Model to Data

Ratcliff and Tuerlinckx (2002) evaluated methods for fitting the diffusion model to experimental data and concluded that the chi-square method provides the best balance between the accuracy with which parameter values are recovered and robustness to contaminant RTs.

The maximum likelihood method takes each individual RT and computes the probability density that it occurred, $f(t)$, using the parameters of the diffusion model. The product of the probabilities is computed by multiplying across the values of $f(t)$ for each RT (in practice, the sum of the logarithms is used), and the parameters of the model are adjusted using a general simplex minimization routine (Nelder & Mead, 1965) to maximize the product (or sum of logs) of the probability densities.

In the chi-square method, the empirical quantile RTs for correct and error responses in each condition of an experiment are computed (most often, the .1, .3, .5, .7, and .9 quantiles). Then, parameter values are adjusted until the model correctly generates the cumulative probability of a response for each quantile RT. Specifically, subtracting the cumulative probability for each successive quantile from that for the next higher quantile gives the proportion of responses between adjacent quantiles. For the chi-square computation, these are the expected values, to be compared with the observed proportions of responses between the quantiles (i.e., the proportions between 0, .1, .3, .5, .7, .9, and 1.0, which are .1, .2, .2, .2, .2, and .1) multiplied by the number of observations. Summing over $(\text{Observed} - \text{Expected})^2/\text{Expected}$ for both correct and error responses gives a single chi-square value. This value is minimized with a general simplex minimization routine (Nelder & Mead, 1965) that adjusts the parameters of the model until the smallest possible chi-square value is obtained (Ratcliff & Tuerlinckx, 2002).

(Manuscript received July 10, 2006;
revision accepted for publication January 24, 2008.)