Psychonomic Bulletin & Review
2007, 14 (1), 3-22

THEORETICAL AND REVIEW ARTICLES

An EZ-diffusion model for response time and accuracy

ERIC-JAN WAGENMAKERS, HAN L. J. VAN DER MAAS, AND RAOUL P. P. P. GRASMAN
University of Amsterdam, Amsterdam, The Netherlands

The EZ-diffusion model for two-choice response time tasks takes mean response time, the variance of response time, and response accuracy as inputs. The model transforms these data via three simple equations to produce unique values for the quality of information, response conservativeness, and nondecision time. This transformation of observed data in terms of unobserved variables addresses the speed–accuracy trade-off and allows an unambiguous quantification of performance differences in two-choice response time tasks. The EZ-diffusion model can be applied to data-sparse situations to facilitate individual subject analysis. We studied the performance of the EZ-diffusion model in terms of parameter recovery and robustness against misspecification by using Monte Carlo simulations. The EZ model was also applied to a real-world data set.

For a two-choice response time (RT) task, the observed variables are response speed and response accuracy. In experimental psychology, inference usually concerns the mean response time for correct decisions (i.e., MRT) and the proportion of correct decisions (i.e., P_c). The immediate problem is that MRT and P_c are in a trade-off relationship: Participants can respond faster, and hence decrease MRT, at the expense of making more errors, thereby decreasing P_c (see, e.g., Pachella, 1974; Schouten & Bekker, 1967; Wickelgren, 1977). This so-called *speed–accuracy trade-off* has for a long time bedeviled the field. Consider 2 participants in an experiment, Amy and Rich. Amy's and Rich's performance is summarized by $MRT = 0.422$ sec, $P_c = .881$, and $MRT = 0.467$ sec, $P_c = .953$, respectively. Amy responds faster than Rich, but she also commits more errors. Thus, it could be that Amy and Rich have the same ability, but Amy risks making more mistakes. It could also be that Amy's ability is higher than that of Rich, or vice versa. If we only consider MRT and P_c, there appears to be no way to tell which of these three possibilities is in fact true.

Now consider George, whose performance is characterized by $MRT = 0.517$ sec, $P_c = .953$. George responds more slowly than Rich, whereas their error rates are identical. An explanation solely in terms of the speed–accuracy trade-off cannot account for this pattern of results, and therefore most researchers would confidently conclude that Rich performs better than George. Unfortunately, if we only consider MRT and P_c, it is impossible to go beyond these conclusions in terms of ordinal relations and quantify *how much* better Rich does than George. Note that the same arguments would hold if the example above had been in terms of 1 participant who responds in three different experimental conditions presented in three separate blocks of trials. In this case, comparison of performance across the different conditions is complicated by the fact that task performance may be simultaneously influenced by task difficulty and response conservativeness.

In sum, both MRT and P_c provide valuable information about task difficulty or subject ability, but neither of these variables can be considered in isolation. When MRT and P_c are considered simultaneously, however, it is not clear how to weigh their relative contributions to arrive at a single index that quantifies subject ability or task difficulty.

A way out of this conundrum is to use cognitive process models to estimate the unobserved variables assumed to underlie performance in the task at hand. The field of research that uses cognitive models for measurement has been termed *cognitive psychometrics* (Batchelder, 1998; Batchelder & Riefer, 1999; Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002), and similar approaches in other paradigms have included those of Busemeyer and Stout (2002); Stout, Busemeyer, Lin, Grant, and Bonson (2004); and Zaki and Nosofsky (2001). Here, the focus is on the diffusion model for two-choice RT tasks (see, e.g., Ratcliff, 1978). In the diffusion model, the three most important unobserved variables are the *quality of information*, *response conservativeness*, and *nondecision time*. A statistical analysis of these unobserved variables is not only immune to speed–accuracy trade-offs, but also affords an unambiguous quantification of performance differences. This article introduces the EZ-diffusion model,

E.-J. Wagenmakers, ewagenmakers@fmg.uva.nl

Table 1
Performance of 4 Hypothetical Participants in a Two-Alternative Forced Choice Task

Participant	RT Mean (sec)	RT Variance	P_c
George	0.517	0.024	.953
Rich	0.467	0.024	.953
Amy	0.422	0.009	.881
Mark	0.372	0.009	.881

Note—Which participant did best? RT denotes response time, and P_c the proportion of correct responses. See the text for details.

a simplified version of the diffusion model that is able to uniquely determine these three important unobserved variables from just three observed quantities: *MRT*, P_c, and the variance of response times for correct decisions (*VRT*). Our mathematical analysis will show that *VRT* is much more informative with respect to subject ability or task difficulty than is *MRT*, echoing recent empirical insights in the aging literature and elsewhere (e.g., Hultsch, MacDonald, & Dixon, 2002; Li, 2002; MacDonald, Hultsch, & Dixon, 2003; Shammi, Bosman, & Stuss, 1998).

An important practical advantage of the EZ-diffusion model is that it does not require a parameter fitting routine (cf. signal detection theory). Also, the EZ-diffusion model can be applied to common experimental setups in which each participant contributes only a moderate amount of data and error rate is low (i.e., 5%–10%).

The outline of this article is as follows. The next section briefly discusses the methods of analysis that are currently standard in the field. Then we briefly describe the "full" Ratcliff diffusion model and introduce the simplified, "EZ" version. Subsequent sections detail the performance of the EZ method in terms of parameter recovery and robustness against misspecification. Then we use a real-world data set to compare parameter estimates for the EZ model against those for the Ratcliff diffusion model. We conclude by stressing the practicality of the present approach and by acknowledging the potential dangers of blindly applying the EZ model to situations in which its assumptions do not hold.

The Standard Analysis of Two-Choice RT Tasks and Its Limitation

For many decades, the analysis of data from two-choice RT tasks has remained unchanged. The standard analysis separately considers *MRT* and P_c. Specifically, one ANOVA is performed for *MRT* and a second for P_c. The standard analysis is simple but crude, and it can be improved in various ways. For instance, Rouder, Lu, Speckman, Sun, and Jiang (2005) recently introduced a Bayesian hierarchical model of Weibull distributions that bases inference not just on *MRT*, but on the entire RT distribution for correct decisions. Similar sophistications (e.g., hierarchical logistic regression) can be proposed for the analysis of P_c.

Regardless of the statistical sophistication that the standard method may undergo, the general framework fails to address the core problem of the two-alternative RT task. This is the problem of how to combine response speed and accuracy in a single index that reflects subject ability or task difficulty.

To highlight the limitations of the standard method of inference, consider the performance of our hypothetical participants in a two-alternative RT task, as shown in Table 1. Assume that the imaginary experiment involves very many trials, so that measurement error is negligible. The standard analysis method is perfectly able to rank-order the participants according to either *MRT* (i.e., 1. Mark, 2. Amy, 3. Rich, 4. George), P_c (i.e., 1&2. George & Rich, 3&4. Mark & Amy), or *VRT* (i.e., 1&2. Mark & Amy, 3&4. George & Rich). However, the standard method cannot rank-order the 4 participants on "ability," since this requires response speed and response accuracy to be combined in some unspecified manner. This also means that the standard method cannot inform us as to how much better or worse one participant performs than another.

The Ratcliff Diffusion Model

A solution to the problem of how to combine response speed and response accuracy is to analyze the data in terms of the parameters of a cognitive model such as a diffusion model. In a diffusion model, illustrated in Figure 1, noisy accumulation of information drives a decision process that terminates when the accumulated evidence in favor of one or the other response alternative exceeds threshold. Thus, the diffusion model is a continuous-time, continuous-state random-walk sequential sampling model (see Laming, 1968; Link, 1992; Link & Heath, 1975; Ratcliff, 1978; Stone, 1960). The reader is referred to Luce (1986), Ratcliff (2002), Ratcliff and Smith (2004), and Townsend and Ashby (1983) for detailed accounts of the diffusion model; to Gardiner (2004), Honerkamp (1994), and Smith (2000) for mathematical foundations; and to Diederich and Busemeyer (2003), Ratcliff and Tuerlinckx (2002), Tuerlinckx (2004), and Voss, Rothermund, and Voss (2004) for a discussion of several methods to fit the model to data.

For concreteness, our focus is on the diffusion model as it applies to the lexical decision task (Ratcliff, Gomez, & McKoon, 2004). In lexical decision, the participant is presented with a letter string that needs to be classified either as an English word (e.g., *zebra*) or as a nonword (e.g., *drapa*). The diffusion model has also been successfully applied to many other two-choice RT paradigms, including short- and long-term recognition memory tasks, same/different letter string matching, numerosity judgments, visual-scanning tasks, brightness discrimination, and letter discrimination (see, e.g., Ratcliff, 1978, 1981, 2002; Ratcliff & Rouder, 1998, 2000; Ratcliff, Van Zandt, & McKoon, 1999). In all these applications, the diffusion model provided a close fit to response accuracy and the observed RT distributions for both correct and error responses.

In the application of the diffusion model to lexical decision, presentation of a word stimulus will generally lead to the accumulation of evidence that supports the correct "word" response, as in the two examples shown in Figure 1. In the model, easy-to-classify letter strings have relatively high absolute drift rate values; that is, these letter strings are associated with relatively high signal-to-noise ratios in the evidence accumulation process. Drift rate ξ is defined

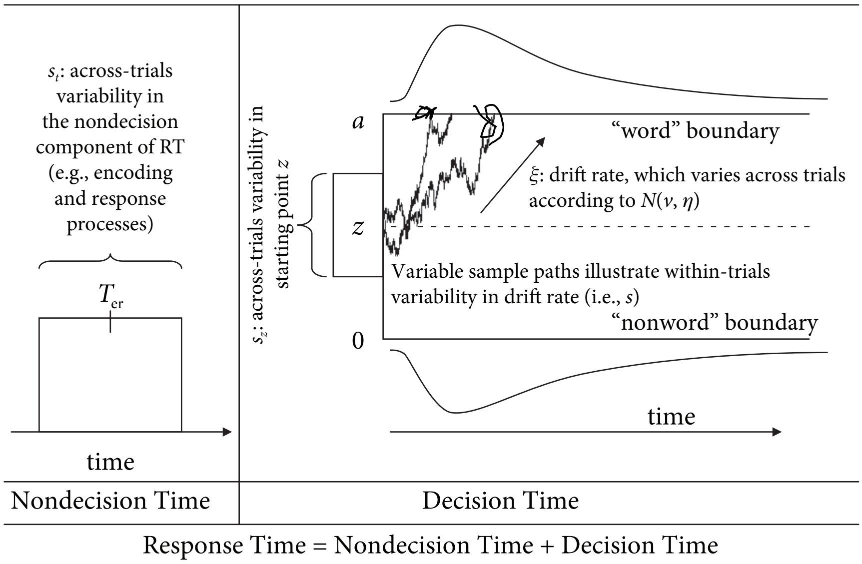

Figure 1. Diffusion model account of evidence accumulation in the lexical decision task (see Ratcliff et al., 2004).

on the real line; $\xi > 0$ and $\xi < 0$ lead to evidence accumulation consistent with a "word" or a "nonword" response, respectively. The case of $\xi = 0$ corresponds to a process that, at each point in time, is equally likely to move upward as it is to move downward. Drift rate is assumed to vary over trials according to $\xi \sim N(v, \eta)$. Because drift rate quantifies the deterministic component of the noisy information accumulation process, it can be interpreted as an index for the signal-to-noise ratio of the information processing system. Therefore, drift rate is an excellent candidate for a measure that combines respond speed and response accuracy to quantify subject ability or task difficulty.

The stochastic, nonsystematic component of the information accumulation process on each trial is quantified by s. The factor $s^2\,dt$ is the variance of the change in the accumulated information for a small time interval dt (Cox & Miller, 1970, p. 208). The s parameter is a scaling parameter, which means that if s doubles, other parameters in the model can be doubled to obtain exactly the same result. Thus, the choice of a specific value for $s > 0$ is arbitrary; in practice, s is usually set to 0.1, and we adhere to this convention throughout the article. Two further important parameters are the boundary separation a and the starting point z. The boundary separation parameter a is especially important here, because large values of a indicate the presence of a conservative response criterion: When a is large, the system requires more discriminative information before deciding on one or the other response alternative. A conservative response criterion results in long response times, but also in highly accurate performance, since with large a it is unlikely that the incorrect boundary will be reached by chance fluctuations. Therefore, in the diffusion model, one of the main mechanisms by which speed–accuracy trade-off phenomena arise is through changes in a.

The a priori bias against one or the other response alternative is given by z. As with drift rate, the exact location of z may fluctuate from trial to trial. This fluctuation is quantified by a uniform distribution with range s_z. As shown later, in most applications z is estimated to be about equidistant from both response boundaries (i.e., $z \approx a/2$). Finally, the diffusion model captures the nondecision component of RT by a parameter T_{er} that varies over trials according to a uniform distribution with range s_t. As is often assumed in RT modeling, the total RT is a sum of the nondecision and decision components of processing (Luce, 1986):

$$\mathrm{RT} = \mathrm{DT} + T_{er}, \tag{1}$$

where DT denotes decision time.

In sum, the Ratcliff diffusion model estimates the following seven parameters:[1]

1. Mean drift rate (v).
2. Across-trials variability in drift rate (η).
3. Boundary separation (a).
4. Mean starting point (z).
5. Across-trials range in starting point (s_z).
6. Mean of the nondecision component of processing (T_{er}).
7. Across-trials range in the nondecision component of processing (s_t).

In theory, these seven parameters could be estimated separately for each experimental condition. In practice, however, only parameters that are believed to be affected by the experimental manipulation are free to vary between conditions.

In order to provide some perspective regarding the ranges of parameter values that may be expected when fitting the Ratcliff diffusion model to data, Figure 2 provides a visual overview of the best-fitting parameter val-

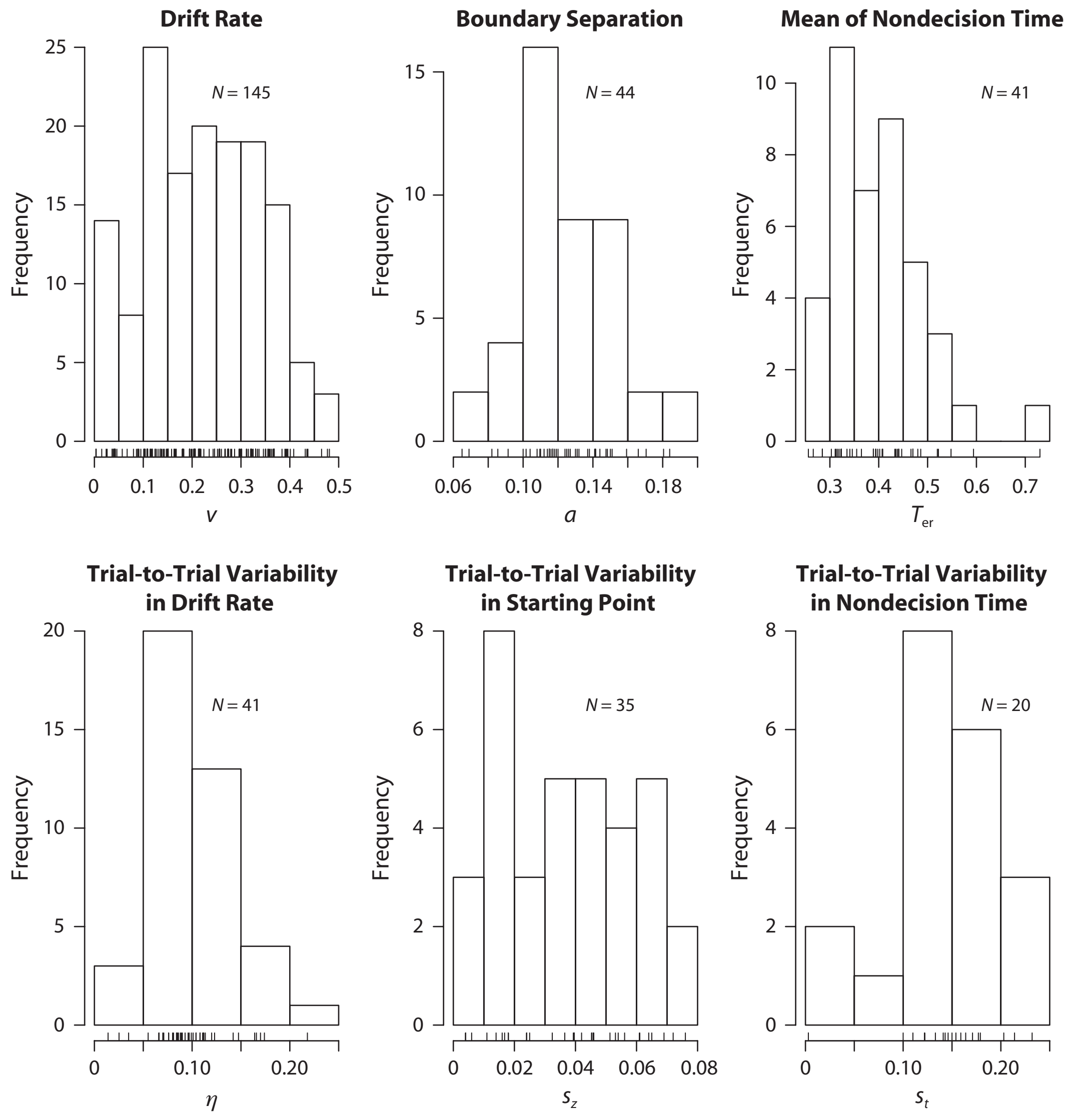

Figure 2. Best-fitting diffusion model parameter values as encountered in previous research. The top left panel plots the absolute values of drift rates (i.e., negative drift rates have been multiplied by −1). The scaling parameter *s* is always fixed at 0.1.

ues encountered in previous experiments (i.e., Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Rouder, 2000; Ratcliff & Smith, 2004; Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2001, 2004; Ratcliff et al., 1999; Van Zandt, Colonius, & Proctor, 2000; Voss et al., 2004). These experiments used tasks such as lexical decision, letter identification, asterisks discrimination, recognition memory, and color discrimination. Studies that manipulated starting point were excluded from consideration. Whenever there was a choice, we selected parameter values estimated from averaged data.[2] Almost all experiments vary task difficulty (i.e., drift rate in the model), and this is the reason why the top left panel contains relatively many values—when a manipulation is thought to affect drift rate, only this parameter is free to vary across conditions. The bottom right panel plots the best-fitting values for the s_t parameter. It represents relatively few experiments because this parameter has been recently added to the diffusion model. Figure 3 shows the relation between boundary separation and starting point as obtained in earlier experiments. The solid line has a slope of 2. Figure 3 confirms the earlier assertion that in many applications, $z \approx a/2$.

The data needed to fit the Ratcliff diffusion model are error rate and RT distributions for correct and error responses. As mentioned earlier, participants usually do not commit very many errors: In most tasks, error rate is lower than 10%. This means that it may take a substantial number of trials to accurately estimate the entire RT distribution for error responses. On the basis of prior experience with the model, a rule of thumb is that about 10 error RTs are needed in order to estimate the error RT distribution with an acceptable degree of reliability. This means that with an error rate of, say, 5%, each experimental condition should contain about 200 observations.

The model is then fit to the data using one of several methods (see, e.g., Ratcliff & Tuerlinckx, 2002). Each method uses the facts that in the diffusion model, the probability of an error (P_e) is given by

$$P_e = 1 - P_c = \frac{\exp\left(-\frac{2av}{s^2}\right) - \exp\left(-\frac{2zv}{s^2}\right)}{\exp\left(-\frac{2av}{s^2}\right) - 1}, \tag{2}$$

and the probability of an error response before time t is given by Equation 3 at the bottom of this page (Cox & Miller, 1970), where k indexes the infinite series and a, z, ξ, and T_{er} are free parameters. As $t \rightarrow \infty$, the part that involves the infinite sum goes to zero, and what remains is simply the probability of an error response. Thus, Equation 3 computes the defective distribution (see, e.g., Ratcliff & Tuerlinckx, 2002). To obtain the equation that gives the probability of a correct response before time t, z and ξ should be replaced by $a - z$ and $-\xi$, respectively.

Although Equation 3 may look daunting,[3] the real problem in fitting the diffusion model is in the fact that parameters T_{er}, z, and ξ vary across trials. Finding the best-fitting values for the across-trials variability parameters s_t, s_z, and η necessitates the use of time-consuming numerical integration procedures. The reason that mathematical psychologists use such a complicated method is the substantial payoff involved. The Ratcliff diffusion model provides a description of response time that is extremely detailed. Perhaps more important, however, is the fact that the parameter values of the model can provide insights that standard, more superficial methods of analysis cannot.

For instance, in an application of the diffusion model to aging (Ratcliff et al., 2001), it was found that in an asterisks discrimination task, older participants responded more slowly but also a little more accurately than the younger participants. The diffusion model was fitted to the data, and the resulting parameter estimates indicated that the parameter that varied between the different age groups was boundary separation a (and T_{er}, the nondecision RT component, which was about 50 msec longer for older adults), whereas mean drift rate v remained fairly constant—if anything, drift rate was a little higher for the group of older participants. This analysis supports the notion that in this particular task, the observed differences in performance arose because the older adults adopted more conservative response criteria than did the younger participants. Such detailed and quantitative conclusions could not be based on a standard ANOVA on the RTs and error rates (see also Oberauer, 2005; Voss et al., 2004).

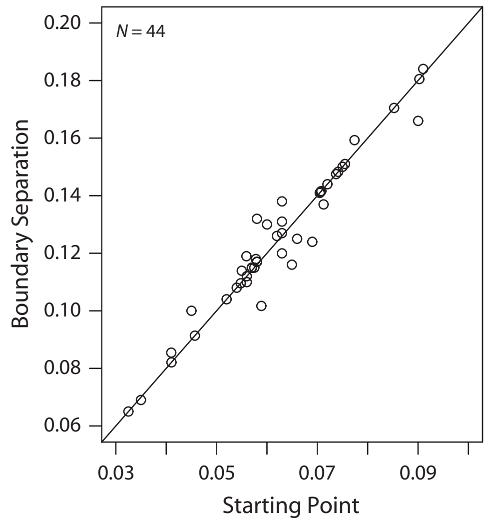

Figure 3. The relationship between starting point and boundary separation as encountered in previous research. The solid line has a slope of 2, suggesting that in many situations the starting point is about equidistant from the two response boundaries.

THE EZ-DIFFUSION MODEL

For a wide range of two-alternative forced choice tasks, the Ratcliff diffusion model provides a principled and seemingly satisfactory solution to the speed-versus-accuracy dilemma that plagues standard methods of analysis. This raises the question as to why the diffusion model is not standardly applied as a psychometric analysis tool. One of the answers is that the Ratcliff diffusion model requires the *entire* RT distribution as input; critically, this includes the RT distribution for incorrect decisions. In many experiments, participants commit few errors overall, and it may take very many trials to obtain an accurate estimate of the error RT distribution. Therefore, in most practical settings it is unclear whether or not the Ratcliff diffusion model can be applied. 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 values may become prominent.

Another important reason why the diffusion model is not used more often in empirical studies is the complexity of the parameter-fitting procedure (see Diederich & Busemeyer, 2003; Ratcliff & Tuerlinckx, 2002; Tuerlinckx,

$$\Pr\left(\text{error}, T \le t\right) = P_e - \frac{\pi s^2}{a^2}\exp\left(\frac{-z\xi}{s^2}\right)\sum_{k=1}^{\infty}\frac{2k\sin\left(\frac{\pi k z}{a}\right)\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2 k^2 s^2}{a^2}\right)\left(t - T_{er}\right)\right]}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2 k^2 s^2}{a^2}\right)} \tag{3}$$

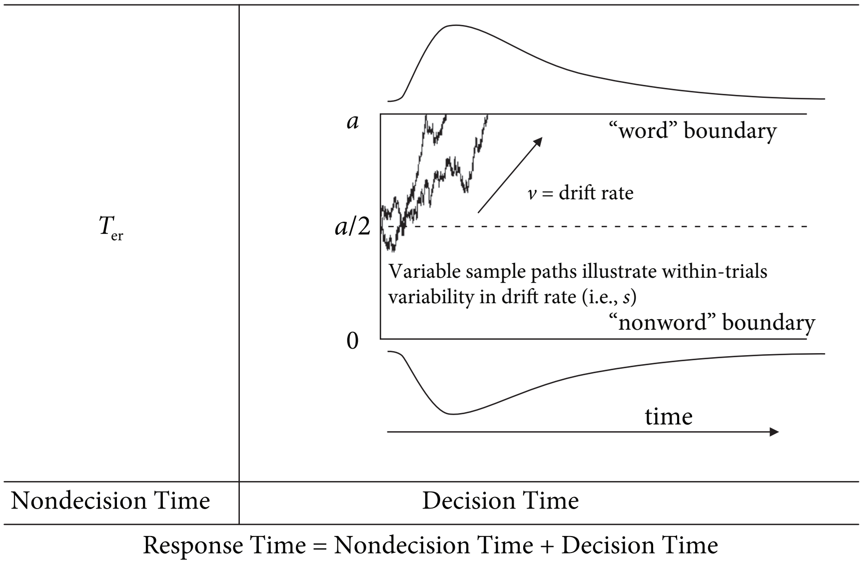

Figure 4. The EZ-diffusion model.

2004). Many experimental psychologists, even those with a firm background in mathematics and computer programming, will find the amount of effort required to fit the Ratcliff diffusion model rather prohibitive.

The EZ-diffusion model constitutes an attempt to popularize a diffusion model analysis of two-alternative forced choice tasks. In order to achieve this goal, we have considerably simplified the Ratcliff diffusion model. These simplifications are warranted by the fact that the aim of the EZ model is much more modest than that of the Ratcliff model. The EZ model tries to determine only the most psychologically relevant parameters of the Ratcliff model: drift rate v (i.e., quality of information), boundary separation a (i.e., response conservativeness), and nondecision time T_{er}. The EZ model does not seek to address the issue of RT distributions, especially not for error responses. Thus, the price that has to be paid for the simplification of the diffusion model is that it no longer provides a very detailed account of the observed behavior, but instead operates at a more macroscopic level. Of course, with few data, this may be the only available option. We will return to this issue in the General Discussion section.

The first simplification is that the EZ-diffusion model does not allow across-trials variability in parameters. This means that s_t, s_z, and η are effectively removed from the model. The effect of s_t—that is, the across-trials variability in T_{er}—is usually not very pronounced (see Ratcliff & Tuerlinckx, 2002). The effect of s_z—that is, across-trials variability in starting point—allows the model to handle error responses that are on average faster than correct responses. The effect of η—that is, across-trials variability in drift rate—is to produce error responses that are on average slower than correct responses. From the bird's-eye perspective taken by the EZ-diffusion model, these aspects of the data are outside the focus of interest.

The second and final simplification is that the starting point z is assumed to be equidistant from the response boundaries, so that $z = a/2$. As mentioned earlier, in practical applications of the diffusion model this is often found to be approximately true (see Figure 3). For instance, Ratcliff et al. (2001) had participants decide whether a screen with asterisks came from a "high" or "low" distribution. Since the design of the stimulus materials was symmetric, one would not expect participants to be biased toward either the "high" or the "low" response category (Ratcliff et al., 2001, p. 332).

In other experiments, however, biases in starting point are more plausible. Consider a hypothetical situation in which participants have an a priori bias to respond "word" to letter strings presented in a lexical decision task. When such a bias exists, the "vanilla" version of the EZ-diffusion model presented here is inappropriate. Fortunately, there exists an easy check for the presence of bias in the starting point: When participants have a starting point bias that favors the "word" response in a lexical decision task, this means that for word stimuli the correct responses are faster than the error responses, whereas for nonword stimuli the correct responses are slower than the error responses. Such a pattern of results indicates a bias in starting point, and this bias renders the results from an EZ-diffusion model analysis suspect. In the General Discussion, we will discuss an extension of the EZ-diffusion model that can be applied to situations in which the starting point is biased. For now, we will work under the assumption that the starting point is equidistant from the response boundaries—that is, that $z = a/2$.

As will soon be apparent, the simplifications above allow the EZ-diffusion model to determine v, a, and T_{er} without a complicated parameter-fitting exercise. Figure 4 shows the EZ-diffusion model and its streamlined set of parameters.

Before proceeding, we should issue a general disclaimer. Any analysis that involves unobserved variables may lead to misleading results when the hypothesized model radically deviates from reality. This holds for both the EZ-diffusion model and the Ratcliff diffusion model. As an example, classical signal detection theory assumes the distributions for "signal plus noise" and "noise only" to have equal variances. When assumptions such as this one are violated, care must be taken with the interpretation of unobserved variables. Fortunately, almost all studies using the diffusion model have shown that the model provides a good description of the RT distributions (Ratcliff, 2002) and that the specific experimental manipulations have selectively affected the model's parameters in the expected direction (see, e.g., Voss et al., 2004). Nevertheless, as with any statistical procedure, one is generally well advised to check whether the data are consistent with the assumptions of the model. We will revisit this issue several times throughout the article.

Mathematical Derivation

The EZ-diffusion model determines drift rate v, boundary separation a, and nondecision time T_{er} from just *MRT*, *VRT*, and P_c. This is possible because we have three unknowns (v, a, and T_{er}) and also three diffusion model equations (for *MRT*, *VRT*, and P_c). As will be apparent later, *VRT* and P_c uniquely determine the values for v and a, so that *MRT* is necessary only to determine T_{er}. This result contrasts sharply with the popular analysis of RTs, which focuses on *MRT* and ignores *VRT* (but see, e.g., Slifkin & Newell, 1998).

The first equation refers to the probability of a correct response—that is, the probability that the stochastic process first arrives at the correct response boundary. Using the fact that $z = a/2$ in the EZ model, Equation 2 simplifies to

$$P_c = \frac{1}{1+\exp\left(-av/s^2\right)}, \tag{4}$$

which can be rewritten as

$$a = \frac{s^2 \operatorname{logit}\left(P_c\right)}{v}, \tag{5}$$

where

$$\operatorname{logit}\left(P_c\right) \equiv \log\left(\frac{P_c}{1-P_c}\right).$$

The second equation refers to the variance of a symmetrical diffusion process (Wagenmakers, Grasman, & Molenaar, 2005). The variance is given by

$$VRT = \left[\frac{as^2}{2v^3}\right]\frac{2y\exp\left(y\right)-\exp\left(2y\right)+1}{\left[\exp\left(y\right)+1\right]^2}, \tag{6}$$

where $y = -va/s^2$ and $v \neq 0$. If $v = 0$,

$$VRT = \frac{a^4}{24s^4}.$$

Palmer, Huk, and Shadlen (2005) independently derived the same equation in terms of hyperbolic functions. Their equation contains a typographical error, and the correct equation is

$$VRT = z_*\left\{\tanh\left(z_*v_*\right)-z_*v_*\left[\operatorname{sech}\left(z_*v_*\right)\right]^2\right\}\Big/v_*^3,$$

where $v_* = v/s$ and $z_* = z/s$.

Substituting Equation 5 for a in Equation 6 and solving for v yields Equation 7, at the bottom of this page. The sign function returns -1 for all negative numbers and 1 for all positive numbers. Inclusion of the $\operatorname{sign}(P_c - 1/2)$ term ensures that v will take on positive values when $P_c > 1/2$ and negative values when $P_c < 1/2$. Using the variance equation derived by Palmer et al. (2005), Equation 7 can also be written as shown at the top of the next page, where $L \equiv \operatorname{logit}(P_c)$. Equation 7 shows that for fixed accuracy, drift rate v in the EZ-diffusion model is inversely proportional to $VRT^{1/4}$, which is the square root of the standard deviation of the RT distribution. When 2 participants respond at the same level of accuracy, their difference in drift rate comes about solely through their difference in *VRT*.

After v has been determined by Equation 7, this allows a to be determined from Equation 5. At this point, the two key parameters v and a have been determined without any recourse to *MRT*. It turns out that *MRT* is useful only to determine the final parameter of the EZ-diffusion model, T_{er}. Recall that in the EZ-diffusion model, as in the Ratcliff diffusion model, *MRT* contains not just the time to classify the stimulus (i.e., decision time), but also the time to visually encode the stimulus and the time to produce a motor response (i.e., nondecision time T_{er}). That is,

$$MRT = MDT + T_{er}, \tag{8}$$

where *MDT* denotes mean decision time.

Given both v and a, *MDT* can be determined from a third equation, which refers to the mean time until arrival at a response threshold:[4]

$$MDT = \left(\frac{a}{2v}\right)\frac{1-\exp\left(y\right)}{1+\exp\left(y\right)}, \tag{9}$$

where, again, $y = -va/s^2$. Given *MDT*, we can now use Equation 8 to obtain T_{er}. Thus, the foregoing discussion

$$v = \operatorname{sign}\left(P_c - \frac{1}{2}\right)s\left\{\frac{\operatorname{logit}\left(P_c\right)\left[P_c^2\operatorname{logit}\left(P_c\right)-P_c\operatorname{logit}\left(P_c\right)+P_c-\frac{1}{2}\right]}{VRT}\right\}^{\frac{1}{4}} \tag{7}$$

$$v = \operatorname{sign}\left(P_c - \frac{1}{2}\right) s \left[\frac{\left\{ L \tanh\left(\frac{1}{2}L\right) - \frac{1}{2}\left[L \operatorname{sech}\left(\frac{1}{2}L\right)\right]^2 \right\}}{2VRT} \right]^{\frac{1}{4}}$$

shows how the EZ-diffusion model transforms *MRT*, *VRT*, and P_c to v, a, and T_{er} without any parameter fitting; all that is needed to determine the parameters is a straightforward computation. The Appendix contains R code (R Development Core Team, 2004) that implements the EZ-diffusion model.

Conceptual Similarity to Signal Detection Analysis

The EZ-diffusion model is very similar to classical signal detection theory (see, e.g., Green & Swets, 1966) in its aim, scope, and method. Figure 5 highlights these similarities. In fact, the EZ-diffusion model can arguably be considered the response time analogue of signal detection theory.[5]

As can be seen from Figure 5, signal detection theory takes hit rate and false alarm rate as input. As output, it produces unique values for discriminability (d') and bias (β). The statistic d' is a fixed property of the condition or the participant, but β is under the control of the participant. Conclusions regarding participant ability or task difficulty that are based solely on hit rates are suspect, since the participant may change the response threshold β to increase hit rates at the expense of increasing false alarm rates.

The EZ-diffusion model takes *MRT*, *VRT*, and P_c as input. As output, it produces unique values for drift rate (v), boundary separation (a), and nondecision time (T_{er}). The drift rate v is a fixed property of the condition or the participant, but a is under the control of the participant. Conclusions regarding participant ability or task difficulty that are based solely on *MRT* or *VRT* are suspect, since the participant may here change the response threshold a to decrease *MRT* and *VRT* at the expense of decreasing P_c.

PARAMETER RECOVERY FOR THE EZ-DIFFUSION MODEL

This section evaluates performance of the EZ-diffusion model in terms of the accuracy with which the model recovers parameter values used to generate simulated data. The Monte Carlo simulations show that the parameters recovered by the model are relatively close to their true values. The variability of the recovered parameter values is acceptable, and decreases with sample size. Bias (i.e., systematic deviation from the true value) is virtually nonexistent. One of the main reasons why the EZ model is able to recover parameters accurately with only few data

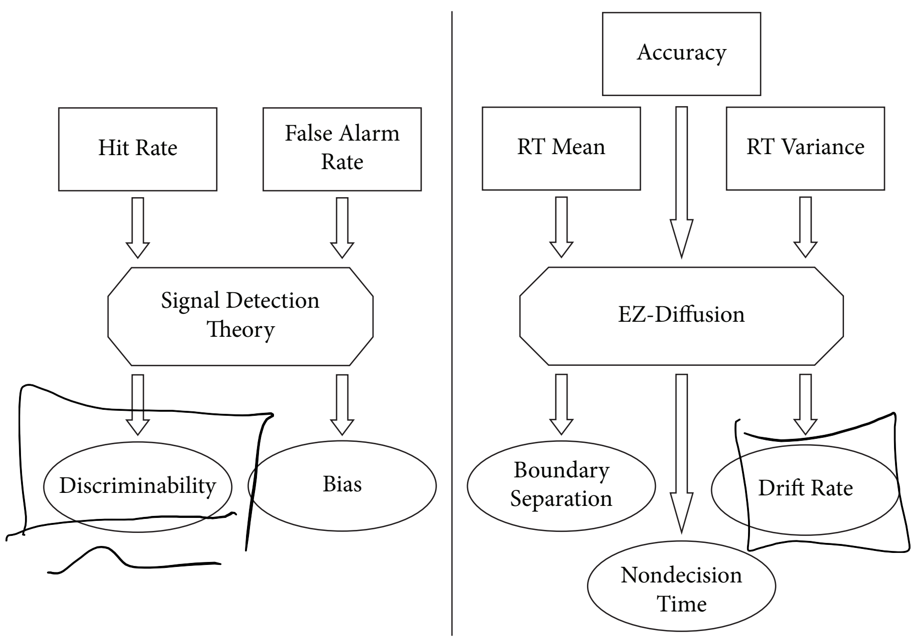

Figure 5. Schematic representation of the similarity between a signal detection analysis and an EZ-diffusion model analysis. The circles at the bottom denote unobserved variables, and the squares at the top denote observed variables. RT, response time.

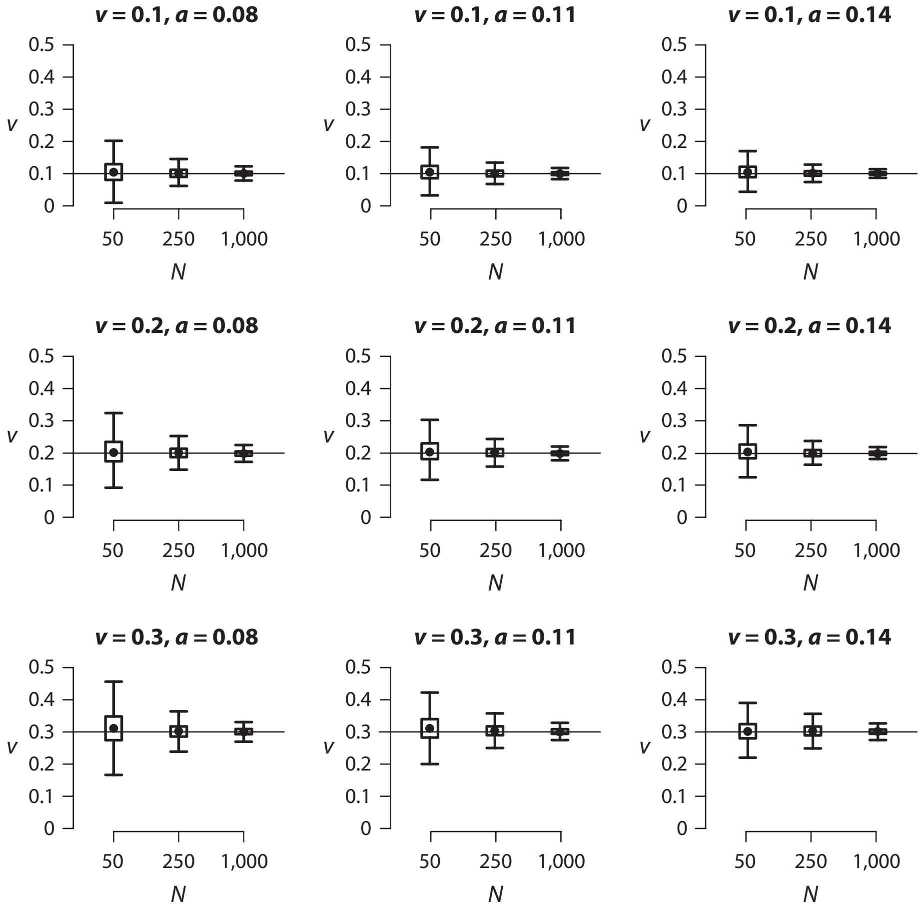

Figure 6. Drift rate parameter recovery for the EZ-diffusion model. Each panel corresponds to a different combination of data-generating parameter values for v and a. The data-generating values for drift rate are indicated by horizontal lines. Each box-plot is based on 1,000 replications.

is that the observed quantities of interest (i.e., *MRT*, *VRT*, and P_c) are estimated relatively efficiently.

In the Monte Carlo simulations reported here, we simulated an experiment with only one condition and a single participant. The experiment had either 50, 250, or 1,000 observations.[6] Also, drift rate v and boundary separation a could each take on one of three values (i.e., $v \in \{0.1, 0.2, 0.3\}$, $a \in \{0.08, 0.11, 0.14\}$). These values were combined to yield $3 \times 3 = 9$ separate sets of parameters that were used to generate simulated data. These parameter values were chosen so as to span a wide range of plausible values (see Wagenmakers et al., 2005). In the simulations, T_{er} was fixed at 0.300. This T_{er} value is arbitrary in the sense that it is an additive constant, the value of which is determined by subtracting the mean decision time from *MRT*. Thus, if T_{er} had been fixed at 0.250, the parameter recovery results would remain the same, save for a constant 50-msec shift. The scaling parameter s was fixed at 0.1, a convention that we adhere to throughout the article.

Next, each of the nine separate parameter combinations was used to generate 1,000 different data sets. For each data set, *MRT*, *VRT*, and P_c were calculated, and the EZ-diffusion model transformations were then applied to yield estimates for v, a, and T_{er}. Note that *MRT* and *VRT* were exclusively based on response times for correct decisions.[7]

When the true values for drift rate v and boundary separation a are relatively large (e.g., $v = 0.3$ and $a = 0.14$), this may result in error-free performance. When $P_c = 1$, Equations 5 and 7 include the undefined term logit(1). The problem is similar to that of applying signal detection theory to a participant who has either a perfect hit rate or a zero false alarm rate—this yields an estimate for d' that is infinite. Several solutions have been proposed to address this issue (see, e.g., Macmillan & Creelman, 2004). Here we chose to apply one of the standard edge-correction methods, replacing $P_c = 1$ with a value that corresponds to one half of an error—that is,

$$P_c = 1 - \frac{1}{2n}.$$

For example, when $n = 50$ and $P_c = 1$, the replacement value for P_c is .99, but when $n = 250$, the replacement value is .998.

Figure 6 shows the results for the parameter recovery simulations with respect to drift rate v. Each panel plots

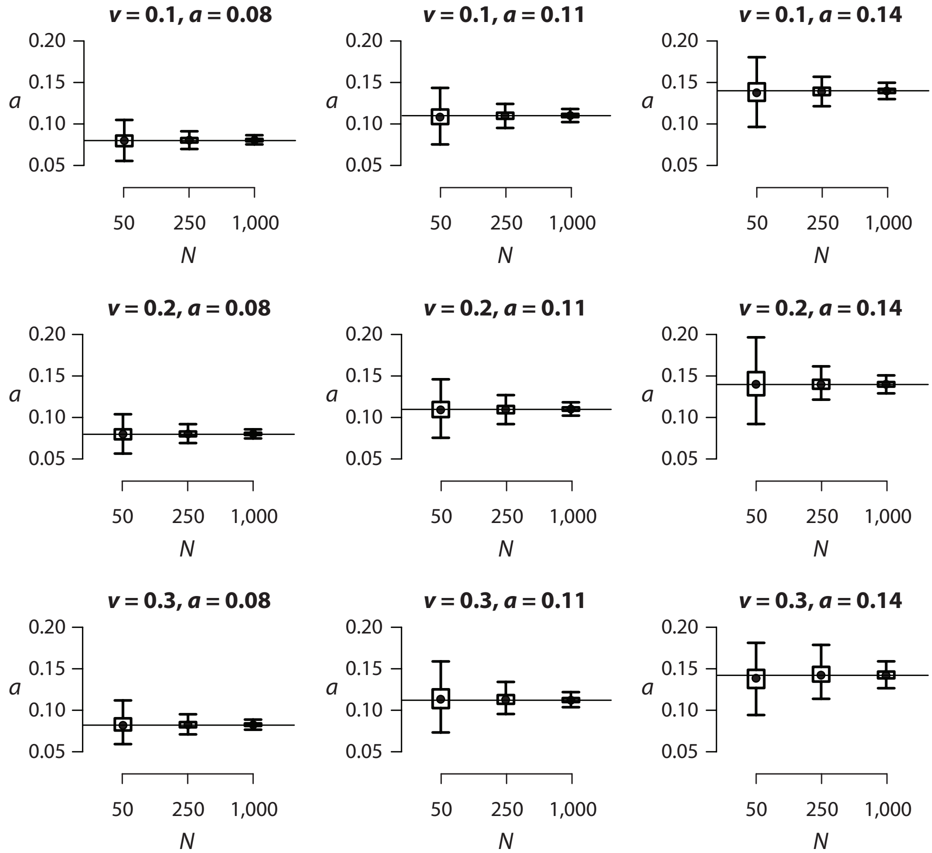

Figure 7. Boundary separation parameter recovery for the EZ-diffusion model. Each panel corresponds to a different combination of data-generating parameter values for *v* and *a*. The data-generating values for boundary separation are indicated by horizontal lines. Each box-plot is based on 1,000 replications.

three box-and-whisker plots, one for each value of $N \in \{50, 250, 1{,}000\}$. A box-and-whisker plot (Tukey, 1977, pp. 39–43) provides an efficient way to summarize an entire distribution, in this case a distribution of recovered parameter values. The box extends from the .25 quantile to the .75 quantile, and the dot in the middle of the box is the .50 quantile (i.e., the median). The whiskers extend to the farthest points that are within 3/2 times the height of the box.

As can be seen from Figure 6, for all panels the median of the recovered parameter values (i.e., the dots in the boxes) tends to coincide with the horizontal line that indicates the generative parameter value. Hence, parameter recovery for v is unbiased. Also note that the whiskers generally extend as far upward as they extend downward, and the dots are in the middle of the boxes. This means that the distributions of recovered parameter values are symmetric. As is to be expected, Figure 6 also clearly shows that the spread of the distributions decreases as N increases. Upon close examination, it appears that recovery of v is subject to more variability when boundary separation a is decreased or drift rate v is increased. Thus, in Figure 6, variability is highest when $v = 0.3$ and $a = 0.08$ (i.e., the leftmost bottom panel), and variability is lowest when $v = 0.1$ and $a = 0.14$ (i.e., the rightmost upper panel).

Figure 7 shows parameter recovery for the boundary separation parameter a. Again, the distributions are symmetric, there is little indication of any bias, and the variability decreases with N. The variability of the distribution of recovered parameter values increases as the true value of a increases—that is, variability increases as we move from the leftward panels to the rightward panels.

Finally, Figure 8 displays the Monte Carlo results for nondecision time T_{er}. Again, there is little evidence of any bias, the distributions appear to be symmetric, and variability decreases markedly with N. The variability for T_{er} increases rather dramatically as boundary separation is increased and drift rate is decreased. Hence, variability in recovery for T_{er} is lowest for the $v = 0.3$, $a = 0.08$ leftmost bottom panel, whereas it is highest for the $v = 0.1$, $a = 0.14$ rightmost top panel. In other words, variability in $T_{er} = MRT - MDT$ increases as MDT (i.e., mean decision time) lengthens.

In sum, the Monte Carlo simulations show that the EZ-diffusion model is able to recover the parameter values for v, a, and T_{er} with virtually no bias. For $N = 50$, the variability in the parameter estimates is considerable. However, it is important to note that this variability is based on a single participant contributing 50 observations. In an experiment with multiple participants, the mean of the in-

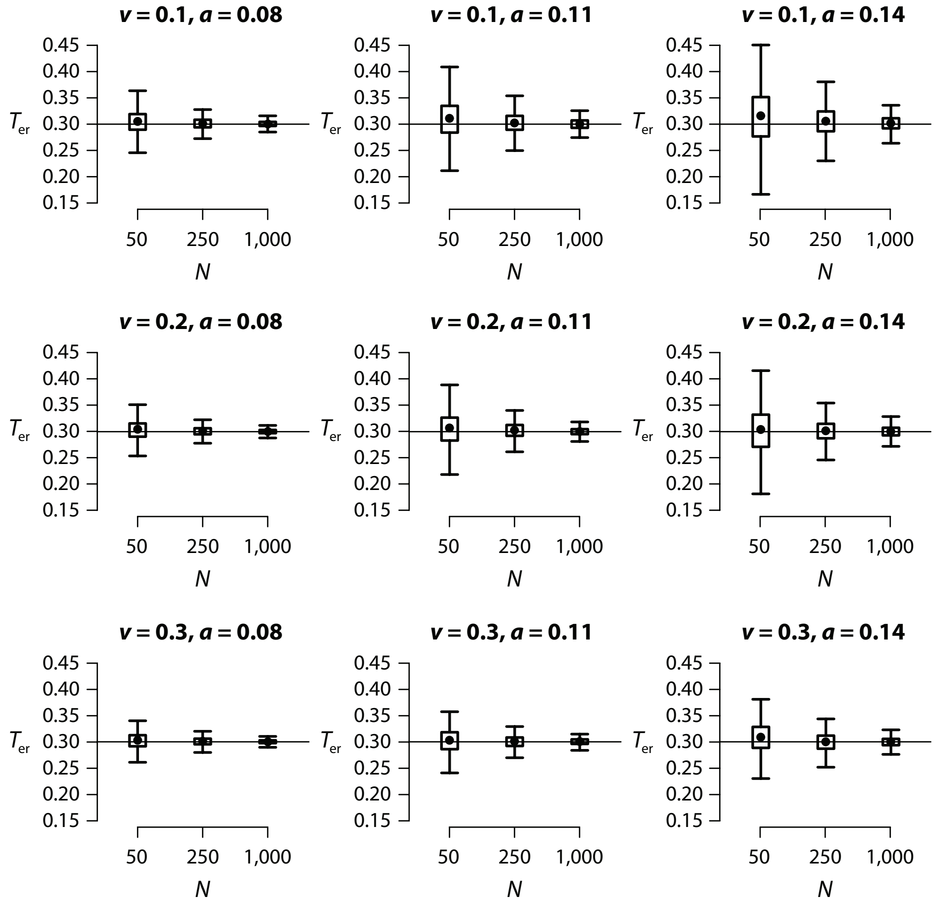

Figure 8. Nondecision time parameter recovery for the EZ-diffusion model. Each panel corresponds to a different combination of data-generating parameter values for v and a. The data-generating value for boundary separation was fixed at T_{er} = 0.300 and is indicated by horizontal lines. Each box-plot is based on 1,000 replications.

dividual parameters will obviously be much less variable than any individual parameter. In practical applications, the variability of the obtained parameter values can always be assessed by sampling the observed data with replacement (i.e., the nonparametric bootstrap; see, e.g., Efron & Tibshirani, 1993). For $N = 250$ and $N = 1{,}000$, the variability is low, even for a single participant.

ROBUSTNESS TO MISSPECIFICATION

The previous section demonstrated that the EZ-diffusion method adequately recovers its parameter values. It is an open question, however, how well the model performs when the data-generating mechanism is different from the one that the EZ-diffusion model assumes. For instance, the EZ-diffusion model assumes that there is no variability across trials in any of the diffusion model parameters. That is, the EZ-diffusion model assumes no across-trials variability in nondecision time (i.e., $s_t = 0$), starting point (i.e., $s_z = 0$), and drift rate (i.e., $\eta = 0$).

In this section, we focus on three situations in which the EZ-diffusion model is "misspecified." First, we consider a data-generating mechanism that has a considerable amount of across-trials variability in nondecision time. Next, we evaluate parameter recovery performance of the EZ-diffusion model in the case in which across-trials variability in drift rate is very high and across-trials variability in starting point is relatively low. Finally, we consider the reverse situation, in which across-trials variability in drift rate is relatively low and across-trials variability in starting point is relatively high. The latter two situations closely resemble those examined by Ratcliff and Tuerlinckx (2002).

In each of the three misspecification analyses reported here, data were generated using three values of drift rate: $v \in \{0.1, 0.2, 0.3\}$. Boundary separation a was fixed at a medium value of 0.11, and nondecision time T_{er} was fixed at 0.300. This yielded three different sets of parameter values. Next, each set of parameter values was used to generate 3,000 data sets: 1,000 data sets with 50 observations each, 1,000 data sets with 250 observations each, and 1,000 data sets with 1,000 observations each. EZ-diffusion parameters were calculated for each data set.

Across-Trials Variability in Nondecision Time

In the first Monte Carlo simulation, the misspecification refers to the presence of across-trials variability in nondecision time. The range of the uniform distribution on T_{er} was set at 0.2 sec, which is at the high end of what is found in empirical research (see, e.g., Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Tuerlinckx, 2002, p. 467; see Figure 2 above, bottom right panel). Figure 9 shows the results of the parameter recovery analysis using box-and-whisker plots. Panels in the top, middle, and bottom rows were generated using $v = 0.1$, $v = 0.2$, and $v = 0.3$, respectively. The horizontal lines indicate the true parameter values.

The panels in the first column of Figure 9 show that the estimation of drift rate remains relatively unaffected by across-trials variability in T_{er}: The values are recovered with little bias, and the variability is not much increased relative to the situation in which $s_t = 0$ (see Figure 6). The panels in the second column show that boundary separation is somewhat overestimated, especially for high values of drift rate. Finally, panels in the third column reveal that nondecision time is somewhat underestimated, and this bias increases with drift rate. Overall, the parameter values are relatively robust against across-trials variability in nondecision time.

Across-Trials Variability in Drift Rate

In the second misspecification analysis, we examined the case of large across-trials variability in drift rate (i.e., normal standard deviation $\eta = 0.16$) and much smaller across-trials variability in starting point (i.e., range of a uniform distribution $s_z = 0.02$). Note that the extent of across-trials variability in η is rather extreme; in empirical work, η is usually smaller (Ratcliff & Tuerlinckx, 2002; see Figure 2 above, bottom left panel).

Figure 10 shows the results. As in the previous figure, panels in the top, middle, and bottom rows were generated using $v = 0.1$, $v = 0.2$, and $v = 0.3$, respectively. It is evident from Figure 10 that the inclusion of a large amount of across-trials variability in drift rate leads to a systematic

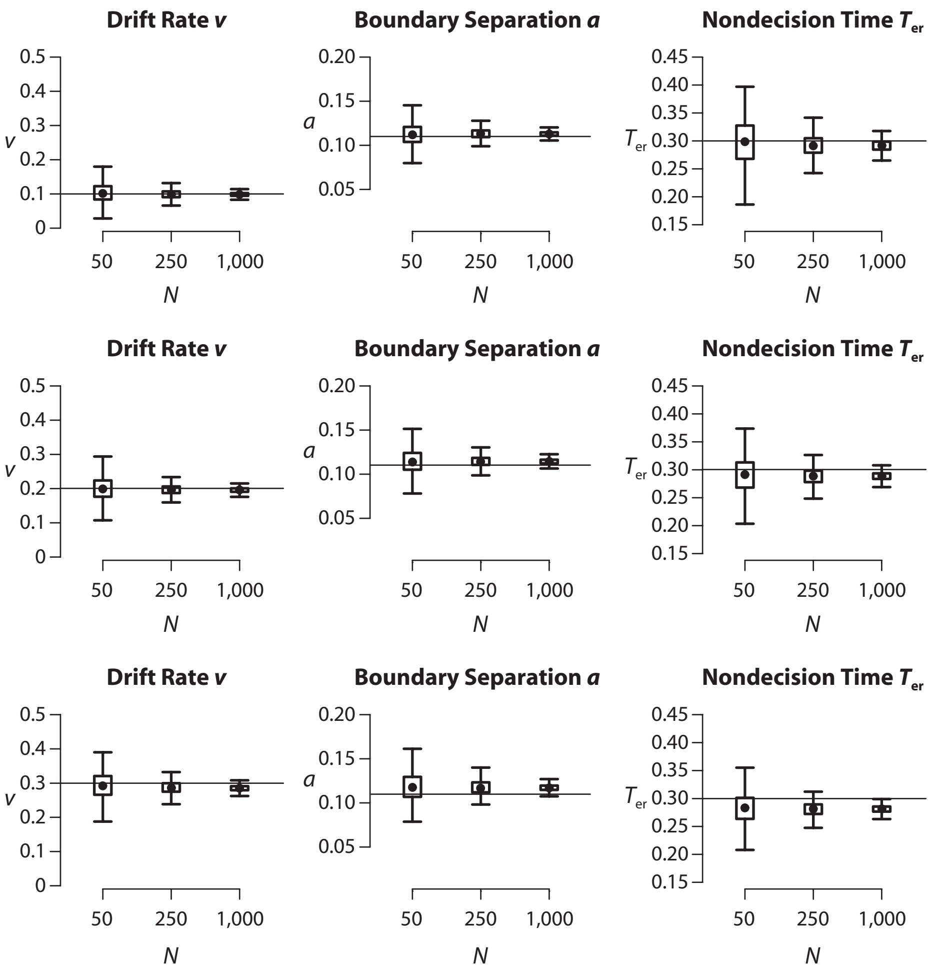

Figure 9. Parameter recovery for the EZ-diffusion model under misspecification, with the data-generating process affected by across-trials variability in nondecision time. The uniform distribution of nondecision time has a range of 0.200 sec, which is at the extreme end of what is observed in practice (Ratcliff & Tuerlinckx, 2002). Boundary separation *a* was fixed at an intermediate value of 0.11, and the mean of the nondecision time T_{er} was fixed at 0.300. Panels in the top, middle, and bottom rows were generated using drift rate values of 0.1, 0.2, and 0.3, respectively. Data-generating parameter values are indicated by horizontal lines. Each box-plot is based on 1,000 replications.

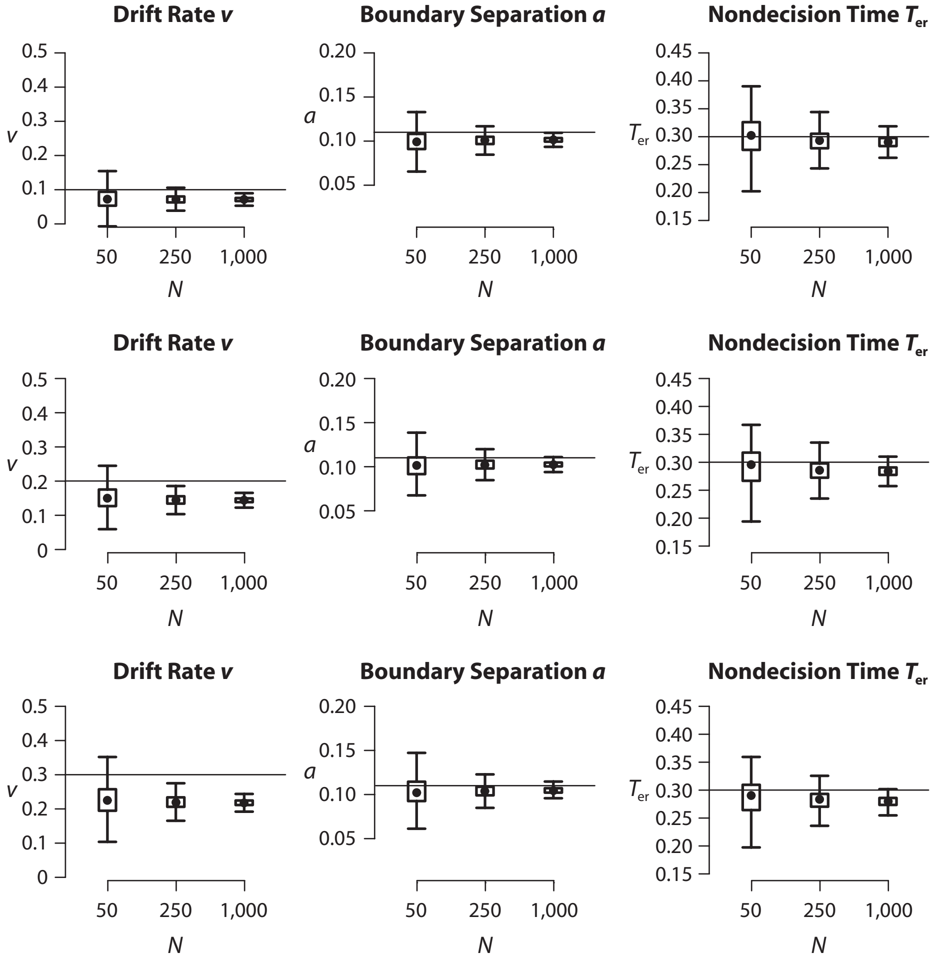

Figure 10. Parameter recovery for the EZ-diffusion model under misspecification with the data-generating process affected by high across-trials variability in drift rate (i.e., η = 0.16), and low across-trials variability in starting point (i.e., s_z = 0.02). The value for η is at the extreme end of what is observed in practice (Ratcliff & Tuerlinckx, 2002). Boundary separation *a* was fixed at an intermediate value of 0.11, and the mean of the nondecision time T_{er} was fixed at 0.300. Panels in the top, middle, and bottom rows were generated using drift rate values of 0.1, 0.2, and 0.3, respectively. Data-generating parameter values are indicated by horizontal lines. Each box-plot is based on 1,000 replications.

underestimation of all three parameters. This bias is not very pronounced for boundary separation (middle column) and nondecision time (right column), but it is quite substantial for drift rate (left column). This drift rate bias is not affected by the number of observations. Although the bias is tolerable for $v = 0.1$, it increases with the estimand, and when $v = 0.3$ the bias is a sizable 0.7. In sum, a substantial amount of across-trials variability in drift rate leads to underestimation of all EZ parameters. This underestimation is particularly pronounced for high values of drift rate.

Across-Trials Variability in Starting Point

A third misspecification analysis was done for the case in which across-trials variability in drift rate is relatively low (i.e., $\eta = 0.08$) whereas across-trials variability in starting point is relatively high (i.e., $s_z = 0.07$; see Figure 2, bottom middle panel). Figure 11 shows that the results are remarkably similar to those of Figure 10: Adding the across-trials variabilities leads to an underestimation of all parameters, and this effect is particularly pronounced for high values of the drift rate parameter (i.e., the leftmost bottom panel). When $v = 0.3$, the bias is a sizeable 0.55.

Overall, the misspecification analyses have shown that for the parameter values under consideration, the EZ-diffusion method is fairly robust to across-trials variability in nondecision time. With large across-trials variabilities in drift rate and starting point, however, all parameters are systematically underestimated. This underestimation is particularly pronounced for high values of drift rate.

These results mean that when the EZ-diffusion model is applied to experimental data, its estimates for drift rate may turn out to be somewhat lower than those of the Ratcliff diffusion model. The empirical data presented later support this assertion: Although the correlations between the EZ parameters and the parameters of the Ratcliff diffusion model are generally quite high, the values for drift

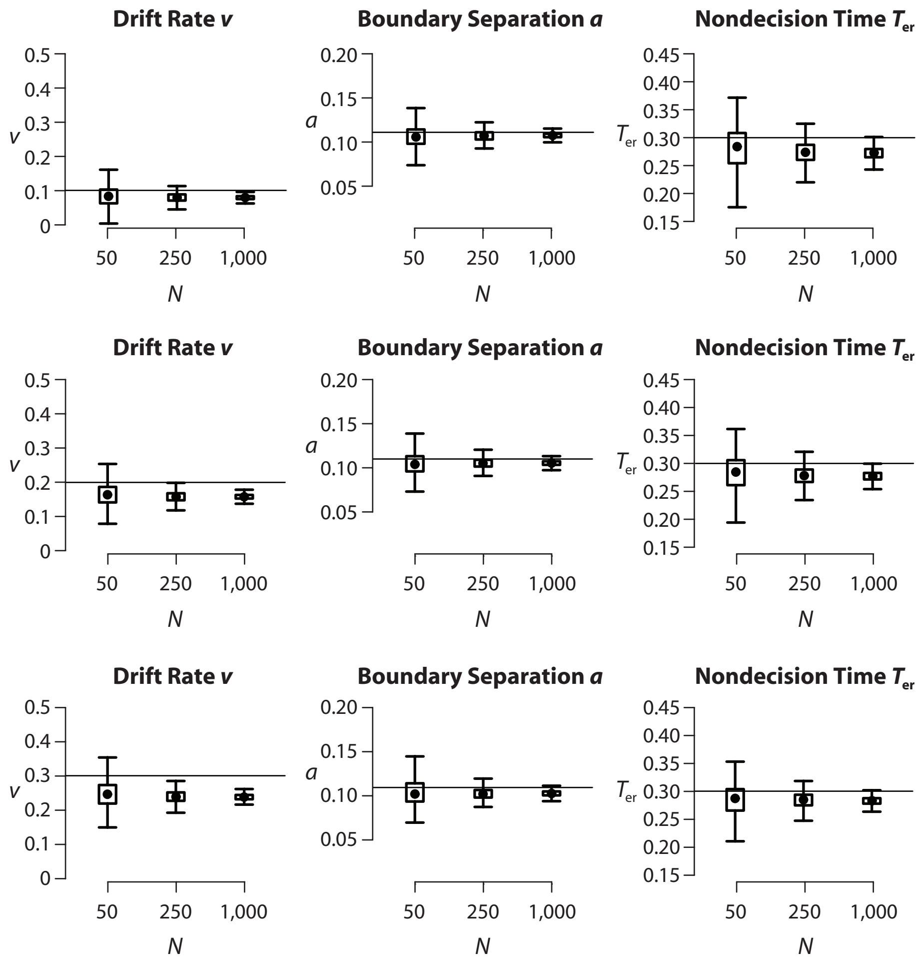

Figure 11. Parameter recovery for the EZ-diffusion model under misspecification, with the data-generating process affected by low across-trials variability in drift rate (i.e., $\eta = 0.08$), and high across-trials variability in starting point (i.e., $s_z = 0.07$). The value for s_z is at the extreme end of what is observed in practice (Ratcliff & Tuerlinckx, 2002). Boundary separation *a* was fixed at an intermediate value of 0.11, and the mean of the nondecision time T_{er} was fixed at 0.300. Panels in the top, middle, and bottom rows were generated using drift rate values of 0.1, 0.2, and 0.3, respectively. Data-generating parameter values are indicated by horizontal lines. Each box-plot is based on 1,000 replications.

rate are systematically lower for the EZ-diffusion model. This effect is magnified for high values of drift rate, as our simulations anticipate.

Three EZ Checks for Misspecification

In practical applications, the assumptions of the EZ-diffusion model may be violated. Depending on the nature and the seriousness of the violation, the results from the EZ-diffusion model should be interpreted with caution or the model should not be applied at all. In order to test whether the EZ-diffusion model is misspecified, we suggest carrying out the following three simple checks. Each check tests a prediction of the model that follows from one of its implicit assumptions.

Check the shape of the RT distributions. The EZ model should be applied only to RT data that show at least some amount of right skew. In addition, the skew should become more pronounced as task difficulty increases. Fortunately, these regularities are present in the wide majority of data sets (see Ratcliff, 2002). If the data are not skewed to the right, or if the skew does not increase with task difficulty, application of the EZ-diffusion model is inappropriate. A statistical test for skewness was proposed by D'Agostino (1970).[8]

Check the relative speed of error responses. As mentioned above, the EZ-diffusion model predicts that the RT distributions of correct and error responses are identical. When the starting point is equidistant from the response boundaries, fast error responses come about through across-trials variability in starting point, and slow error responses come about through across-trials variability in drift rate. Fast or slow errors therefore indicate the presence of across-trials variability in starting point or drift rate, respectively. As shown above, the EZ-diffusion model ignores the across-trials variabilities and this leads to an underestimation of all parameters, in particular drift rate. Standard parametric and nonparametric tests may be used to check whether errors are systematically faster or slower than correct responses.

Check whether the starting point is unbiased. The present version of the EZ-diffusion model assumes that the two stimulus categories in a two-alternative response time task are a priori equally attractive. This means that the starting point z is equidistant from the two response boundaries—that is, $z = a/2$. In many situations, this simplification may be acceptable (see Figure 3). In other situations (e.g., when experimental manipulations include differential payoffs or different presentation rates), the EZ assumption that $z = a/2$ is almost surely violated, and the model should then be applied only with extreme caution. In order to check whether or not the data show evidence of a bias in starting point, one can compare the relative speed of correct and error responses for the different stimulus categories. When participants have an a priori bias that favors Category A over Category B, correct responses should be faster than error responses for Category A stimuli, whereas correct responses should be slower than error responses for Category B stimuli. As a statistical test, one can first determine whether or not stimulus category interacts with response correctness, and then plot the mean RTs to visually judge whether the interaction crosses over in such a way that errors are fast for one stimulus category and slow for the other.

APPLICATION TO AN EXPERIMENT ON PERCEPTUAL DISCRIMINATION

One of the most convincing ways to show that the EZ-diffusion model presents a reasonable alternative to the Ratcliff diffusion model is to compare the parameter estimates for both models on a set of empirical data. Here we consider data from a perceptual discrimination experiment (Meevis, Luth, vom Kothen, Koomen, & Verouden, 2005), to which we fit both the EZ model and the Ratcliff model on a participant-by-participant basis.

The task of each participant was to indicate, as quickly as possible without making errors, which of two vertical line segments was longer. The line segments were presented side by side and were joined by a horizontal line, either at the top or at the bottom. The 100-msec presentation of the line segments was terminated by the presentation of a mask. Task difficulty was manipulated on three levels (i.e., easy, medium, and difficult) by varying the difference in length between the vertical line segments: In the easy, medium, and difficult conditions, the length difference was 2, 4, and 6 mm, respectively.

Eighty-eight university students completed an 18-trial practice block, followed by a total of 1,992 experimental trials in two blocks (i.e., 1,992/3 = 664 trials for each level of difficulty). Twelve participants had an excessive number of fast guesses (i.e., over 100 trials with response times below 250 msec), and these participants were excluded from the analysis. Their exclusion did not affect the qualitative pattern of results. Thus, the EZ-diffusion model and the Ratcliff diffusion model were applied to the data from $N = 76$ participants.[9] The EZ-diffusion model was then used to determine v, a, and T_{er} for each participant and each difficulty level separately, yielding $76 \times 3 = 228$ sets of parameter values. The Ratcliff diffusion model was likewise used to determine v, a, and T_{er}.[10] The EZ-diffusion model parameters were used as starting values for the Ratcliff diffusion model fitting routine.

Figure 12 shows that the EZ parameters correlate quite highly with parameter estimates obtained using the Ratcliff diffusion model. Averaged across all nine panels, the correlation is .867. In the panels that correspond to drift rate and boundary separation, the slope of the best-fitting line is decidedly smaller than 1. This indicates that the EZ-diffusion estimates are lower than those of the Ratcliff diffusion model. For drift rate, this effect is most pronounced for high drift rates, as is evident from the flattening that occurs in the panels corresponding to the easy and medium conditions. As mentioned earlier, this effect may well be due to the fact that the Ratcliff diffusion model has three variability parameters that soak up some of the variance that the EZ-diffusion model attributes to drift rate and boundary separation.

To verify that the implicit assumptions of the EZ-diffusion model had been met, the EZ checks were carried out for all 76 participants and all 3 difficulty levels, resulting in 228 statistical comparisons for each check. The first check used the D'Agostino test for skewness (D'Agostino, 1970) and confirmed that the RT distributions were clearly right-skewed. The results from the second and third checks were more ambiguous. The second check used the ANOVA procedure to test whether correct responses were as fast as error responses. Without any correction for multiple testing and an alpha level of .05, 14 out of 76 participants failed this test for all three levels of difficulty. The majority of the participants failed this test for at least one level of difficulty. For some of the participants, errors were systematically faster than correct responses, and for others errors were systematically slower than correct responses. After the Bonferroni correction was applied and the alpha level consequently reduced to .05/228 = .0002, 6 participants still failed the test for all three levels of difficulty, and 19 failed the test for at least one level of difficulty. These results suggest that there might have been substantial across-trials variability in starting point and drift rate, at least for some of the participants.

The third check used the ANOVA procedure to test whether errors were fast for one stimulus category and slow for the other, since this pattern is indicative of a bias in starting point (i.e., $z \neq a/2$). If the starting point is biased, one would expect the interaction between stimulus category and response correctness to be present for all three difficulty levels. Without any correction for multiple testing and an alpha level of .05, 6 out of 76 participants showed a significant crossover interaction for at least two of the levels of difficulty. Twenty-two participants showed at least one significant crossover interaction. After applying the Bonferroni correction, none of the participants showed the crossover interaction for at least two levels of difficulty, and only 2 out of 76 showed at least one significant crossover interaction. These results suggest that some participants might have had a bias in starting point. Exclusion of the participants that failed the second or third EZ checks did not greatly influence the pattern of correlations.

Table 2
Performance of the 4 Participants From Table 1 in Terms of EZ-Diffusion Model Parameters

Participant	Drift Rate	Boundary Separation	Nondecision Time
George	0.25	0.12	0.300
Rich	0.25	0.12	0.250
Amy	0.25	0.08	0.300
Mark	0.25	0.08	0.250

Note—Participants differed in terms of response conservativeness and nondecision time, but not in terms of efficiency of stimulus processing. See the text for details.

In sum, the parameter values as determined by the EZ-diffusion model correlate highly with those estimated by the diffusion model. Despite this high correlation, the EZ-diffusion model systematically yields estimates of drift rate and boundary separation that are lower than those of the Ratcliff diffusion model. For the drift rate parameter, this effect is most pronounced when drift rate is high.

DISCUSSION

In the context of psychometric testing, Dennis and Evans state that "it is important to recognize that there is no 'magic formula' which will solve the problem of different individuals adopting different speed–accuracy compromises by collapsing the two measures into a single number representing ability" (Dennis & Evans, 1996, p. 123). The aim of the present article was to present just such a formula for the kinds of speeded two-choice tasks that have been popular in experimental psychology for decades. The EZ-diffusion model does not just compute a measure of ability or information uptake (i.e., drift rate), it also yields measures for response conservativeness (i.e., boundary separation) and nondecision time (for approaches with a similar focus, see Balakrishnan, Busemeyer, MacDonald, & Lin, 2002; Palmer et al., 2005; Reeves, Santhi, & Decaro, 2005).

Thus, the EZ-diffusion model transforms the observed variables to three unobserved variables, so that statistical inference can be performed on the latent rather than on the observed variables. The advantages of operating on the level of latent variables is that each variable has a clear psychological interpretation—in contrast, the traditional method of analysis considers both response speed and response accuracy but is at a loss as to how to combine these measures. The conceptual advantages of the EZ-diffusion model are illustrated by Table 2, which shows the latent variables for the data from Table 1 presented at the start of this article.

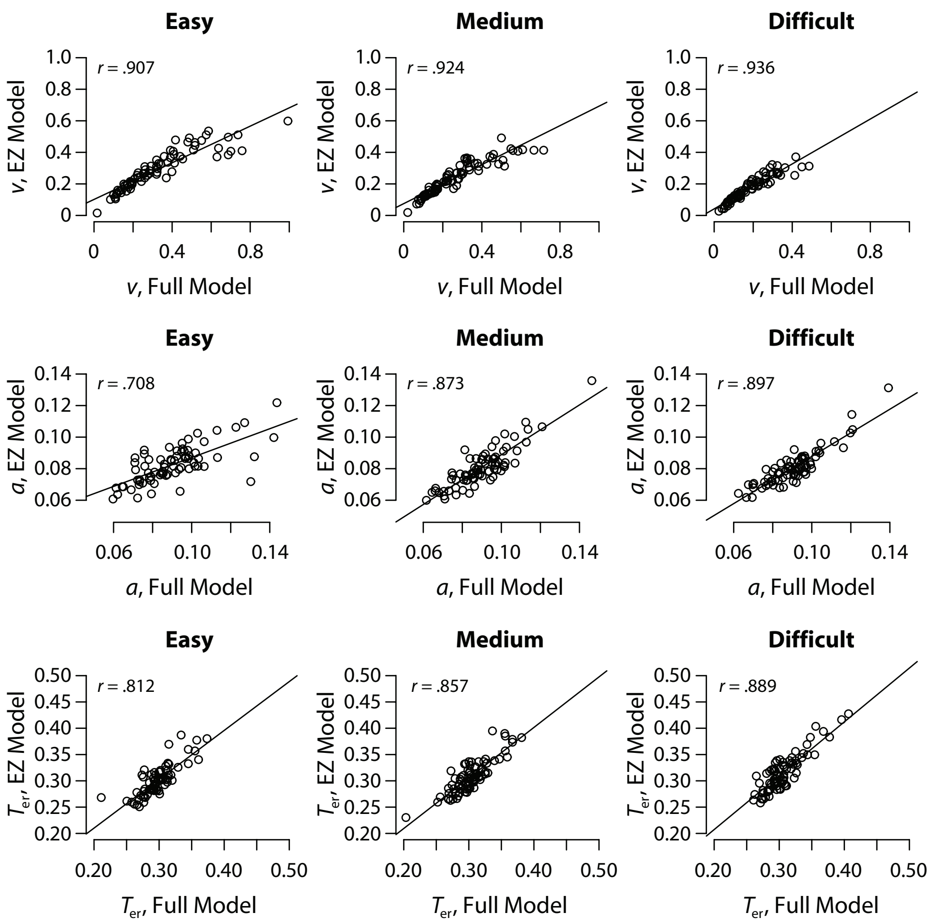

Figure 12. Parameter estimates of the Ratcliff diffusion model and the EZ-diffusion model for a two-choice perceptual discrimination experiment ($N = 76$) featuring three difficulty levels.

From the EZ parameters in Table 2, it is immediately clear that information uptake (i.e., drift rate) is the same for all participants. The reason that George responds relatively slowly is because he is cautious not to make errors (i.e., boundary separation $a = 0.12$) and has a relatively long nondecision time (i.e., $T_{er} = 0.300$). Mark, the fastest responder, is the opposite of George, in that Mark is a risky decision maker (i.e., $a = 0.08$) who has relatively short nondecision time. Amy and Rich differ from each other in that Amy is less cautious than Rich, but Rich has a shorter nondecision time. These kinds of psychologically meaningful conclusions can never be derived by the standard analysis of two-choice tasks.

A Cautionary Note on Transformations and Falsifiability

A considerable practical advantage of the EZ-diffusion model is that it does not require any fitting: The EZ equations simply transform the observed quantities of *MRT*, *VRT*, and P_c to the unobserved quantities of drift rate, boundary separation, and nondecision time. This practical advantage, however, does come at a theoretical cost. That is, the EZ equations will do their job, regardless of whether or not the EZ model is appropriate to the situation at hand. For instance, the data under consideration could be uniformly distributed, left-skewed, or even multimodal. In these cases, it is almost certain that the data do not originate from a diffusion process with absorbing boundaries, as shown in Figure 4.

Despite the fact that the EZ model is not appropriate for, say, multimodal distributions, the EZ transformation will nevertheless return estimated values of drift rate, boundary separation, and nondecision time. Consequently, these estimated values may very well lead to conclusions that are unwarranted. It should always be kept in mind that the EZ-diffusion transformation is only appropriate when the implicit assumptions of the EZ-diffusion model are met. In sum, the EZ-diffusion model cannot be falsified on the basis of a poor fit to the data: It will always produce a perfect fit to the data, since it simply transforms the observed variables to unobserved variables without any loss of information (see Figure 5).

What this means is that some attention should be paid to the underlying assumptions of the EZ-diffusion model when applying it to data. For instance, both the EZ- and Ratcliff diffusion models are currently limited to tasks that require only a single process for their completion. That is, the present model should not be applied to tasks such as the Eriksen flanker task (Eriksen & Eriksen, 1974), in which one process may correspond to information accumulation from the target arrow, and another process may correspond to information accumulation from the distractor arrows. We strongly recommend that the three EZ checks for misspecification mentioned earlier (i.e., check the shape of the RT distributions, check the relative speed of error responses, and check whether the starting point is unbiased) be carried out when the model is applied to data.

Future Directions and Extensions

The EZ-diffusion model described here can be extended in several ways. First and foremost, the current “vanilla” version of the EZ-diffusion model assumes that both stimulus alternatives are equally preferable a priori—that is, that $z = a/2$. However, it is possible to extend the EZ-diffusion model to handle biased starting points—that is, cases for which $z \neq a/2$. Consider again the lexical decision task, and assume that we need to estimate a number of variables: drift rate for word stimuli v_w, drift rate for nonword stimuli v_{nw}, boundary separation a, starting point z, nondecision time for word stimuli $T_{er,w}$, and nondecision time for nonword stimuli $T_{er,nw}$. These six parameters can be obtained by transformation from the six observed variables MRT_w, MRT_{nw}, VRT_w, VRT_{nw}, $P_{c,w}$, and $P_{c,nw}$.

Second, the present version of the EZ-diffusion model does not allow parameters to be constrained across conditions. This may be desirable for several reasons. Consider, for instance, an experiment designed to compare task performance of young adults with that of older adults. The hypothesis that the locus of the aging effect is in the efficiency of information processing corresponds to an EZ-diffusion model in which only drift rate is free to vary between the age groups. A rival hypothesis may entail that the locus of the aging effect is in response conservativeness, and this corresponds to an EZ-diffusion model in which only boundary separation is free to vary between the age groups.

When parameters are constrained across experimental conditions or groups of participants, the number of observed variables becomes larger than the number of unobserved parameters, and this necessitates the use of model fitting. This fitting procedure requires that the lack of fit for *MRT*, *VRT*, and P_c be weighted, for instance by the precision with which these quantities are estimated (i.e., weighted least squares; Seber & Lee, 2003). Once parameters have been constrained and their optimal values determined by the weighted least-squares model-fitting procedure, the model selection issue becomes prominent again: Which model is better, the one in which the effect of age is attributed to differences in information uptake, or the one in which the age effect is due to differences in response conservativeness? For the EZ-diffusion model, an attractive model selection procedure would be to use split-half cross-validation (see, e.g., Browne, 2000): That is, the parameters of the model could be determined by fitting one half of the data set. These particular parameter estimates could then be used to assess the prediction error for the second half of the data set. The model with the lowest prediction error would be preferred.

EZ Diffusion or Ratcliff Diffusion?

The EZ-diffusion model is a considerable simplification of the Ratcliff diffusion model. This is both good and bad. One of the advantages of using a simple model is that the results are more readily interpretable—hence, more easily communicated to other researchers. Another advantage is that simple models are easily implemented. Furthermore, simple models such as the EZ-diffusion model can be applied to very large data sets in a matter of seconds. Finally, simple models are less prone to overfitting (i.e., modeling noise), and may therefore yield relatively low prediction errors to unseen data from the same source (see, e.g., Myung, Forster, & Browne, 2000; Wagenmakers & Waldorp, 2006).

A disadvantage of a simple model such as the EZ model is that it may not capture all aspects of reality that one might consider important. For instance, with the starting point equidistant from the response boundaries and no across-trials variability in drift rate, the diffusion model predicts that the RT distribution for correct responses is identical to the one for error responses. Empirical work has shown that this is not always the case; errors can be systematically faster or systematically slower than correct responses (see, e.g., Ratcliff & Rouder, 1998). In contrast to the EZ-diffusion model, the Ratcliff diffusion model provides an elegant account of the relative speed of errors versus correct responses.

In this context, it is important to realize that the Ratcliff diffusion model is also a simplification of a diffusion process with even more variables. For instance, the current mainstream version of the model (see, e.g., Ratcliff & Tuerlinckx, 2002) falsely assumes the absence of sequential effects (i.e., repetitions vs. alternations of stimuli; see Luce, 1986, pp. 253–271) and serial correlations (see, e.g., Gilden, 2001; but see Wagenmakers, Farrell, & Ratcliff, 2004). Furthermore, the Ratcliff diffusion model does not assume any across-trials variability in boundary separation, despite the fact that it is very unlikely that participants are equally cautious on every trial of an experiment. Finally, the diffusion model does not have a control structure that is able to set, keep track of, and adjust the boundary separation parameter (see Botvinick, Braver, Barch, Carter, & Cohen, 2001; Jones, Cho, Nystrom, Cohen, & Braver, 2002; Vickers & Lee, 1998).

At this point, it is useful to recall George Box's famous adage "All models are wrong, but some are useful" (Box, 1979, p. 202). The EZ-diffusion model is certainly useful in that it estimates the three most important unobserved variables of the Ratcliff diffusion model with minimal demands regarding the amount of data and the level of mathematical sophistication of the researcher.

In sum, the EZ-diffusion model cannot and should not replace the Ratcliff diffusion model, in the same way that the U.S. "EZ" tax forms cannot and should not replace the more elaborate tax forms. The choice of whether to apply the EZ-diffusion model or the Ratcliff diffusion model may therefore be determined to a large extent by the specific aim of the researcher. When the aim is to precisely describe the RT distributions or to study the relation between correct and error response times, the Ratcliff diffusion model is obviously the right choice. When the aim is to address the speed–accuracy trade-off and estimate unobserved variables such as nondecision time, drift rate, and boundary separation, the EZ-diffusion model presents an attractive alternative.

AUTHOR NOTE

This research was funded by a VENI grant from the Dutch Organization for Scientific Research (NWO). Part of this work was presented at the 4th Annual Summer Interdisciplinary Conference, Briançon, France (July 2005), and at the 38th Annual Meeting of the Society for Mathematical Psychology, Memphis, Tennessee (August 2005). We thank Andrew Heathcote and Francis Tuerlinckx for making their diffusion model fitting routines available to us. Correspondence concerning this article may be addressed to E.-J. Wagenmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands (e-mail: ewagenmakers@fmg.uva.nl).

REFERENCES

Balakrishnan, J. D., Busemeyer, J. R., MacDonald, J. A., & Lin, A. (2002). *Dynamic signal detection theory: The next logical step in the evolution of signal detection analysis* (Cognitive Science Tech. Rep. No. 248). Bloomington: Indiana University, Cognitive Science Program.

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment*, **10**, 331-344.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, **6**, 57-86.

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.

Box, G. E. P. (1979). Robustness in scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 201-236). New York: Academic Press.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, **44**, 108-132.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, **14**, 253-262.

Cox, D. R., & Miller, H. D. (1970). *The theory of stochastic processes*. London: Methuen.

D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g_1. *Biometrika*, **57**, 679-681.

Dennis, I., & Evans, J. B. T. (1996). The speed–error trade-off problem in psychometric testing. *British Journal of Psychology*, **87**, 105-129.

Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, **47**, 304-322.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Emerson, P. L. (1970). Simple reaction time with Markovian evolution of Gaussian discriminal processes. *Psychometrika*, **35**, 99-109.

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

Gardiner, C. W. (2004). *Handbook of stochastic methods* (3rd ed.). Berlin: Springer.

Gilden, D. L. (2001). Cognitive emissions of 1/*f* noise. *Psychological Review*, **108**, 33-56.

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

Honerkamp, J. (1994). *Stochastic dynamical systems: Concepts, numerical methods, data analysis* (K. Lindenberg, Trans.). New York: VCH.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journals of Gerontology*, **57B**, P101-P115.

Jones, A. D., Cho, R. Y., Nystrom, L. E., Cohen, J. D., & Braver, T. S. (2002). A computational model of anterior cingulate function in speeded response tasks: Effects of frequency, sequence, and conflict. *Cognitive, Affective, & Behavioral Neuroscience*, **2**, 300-317.

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.

Laming, D. R. J. (1973). *Mathematical psychology*. London: Academic Press.

Li, S.-C. (2002). Connecting the many levels and facets of cognitive aging. *Current Directions in Psychological Science*, **11**, 38-43.

Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Erlbaum.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77-105.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria longitudinal study. *Psychology & Aging*, **18**, 510-523.
MacMillan, N., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
Meevis, M., Luth, I., vom Kothen, L., Koomen, A., & Verouden, J. (2005). *IQ en reactiesnelheid: Een experiment en een wiskundige analyse* (Tech. Rep.). Amsterdam: University of Amsterdam.
Myung, I. J., Forster, M. R., & Browne, M. W. (Eds.) (2000). Model selection [Special issue]. *Journal of Mathematical Psychology*, **44**(1).
Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, **134**, 368-387.
Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41-82). Potomac, MD: Erlbaum.
Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, **5**, 376-404.
Pew, R. W. (1969). The speed–accuracy operating characteristic. *Acta Psychologica*, **30**, 16-26.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, **88**, 552-572.
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., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347-356.
Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 127-140.
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. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory & Language*, **50**, 408-424.
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*, **102**, 261-300.
R Development Core Team (2004). *R: A language and environment for statistical computing*. Vienna: Author.
Reeves, A., Santhi, N., & Decaro, S. (2005). A random-ray model for speed and accuracy in perceptual experiments. *Spatial Vision*, **18**, 73-83.
Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, **14**, 184-201.
Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, **12**, 573-604.
Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, **12**, 195-223.
Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, **27**, 143-153.
Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis* (2nd ed.). Hoboken, NJ: Wiley.
Shammi, P., Bosman, E., & Stuss, D. T. (1998). Aging and variability in performance. *Aging, Neuropsychology, & Cognition*, **5**, 1-13.
Slifkin, A. B., & Newell, K. M. (1998). Is variability in human performance a reflection of system noise? *Current Directions in Psychological Science*, **7**, 170-177.
Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, **44**, 408-463.
Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, **25**, 251-260.
Stout, J. C., Busemeyer, J. R., Lin, A., Grant, S. J., & Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic Bulletin & Review*, **11**, 742-747.
Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and probability density functions in the diffusion model. *Behavior Research Methods, Instruments, & Computers*, **36**, 702-716.
Tukey, J. W. (1977). *Explanatory data analysis*. Reading, MA: Addison-Wesley.
Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, **7**, 208-256.
Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, & Life Sciences*, **2**, 169-194.
Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, **32**, 1206-1220.
Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f^{\alpha}$ noise in human cognition. *Psychonomic Bulletin & Review*, **11**, 579-615.
Wagenmakers, E.-J., Grasman, R. P. P. P., & Molenaar, P. C. M. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology*, **49**, 195-204.
Wagenmakers, E.-J., & Waldorp, L. (Eds.) (2006). Model selection: Theoretical developments and applications [Special issue]. *Journal of Mathematical Psychology*, **50**(2).
Wickelgren, W. A. (1977). Speed–accuracy trade-off and information processing dynamics. *Acta Psychologica*, **41**, 67-85.
Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1022-1041.

NOTES

1. Recently, Ratcliff and Tuerlinckx (2002) proposed parameter number eight, which is the probability of an RT "contaminant." The inclusion of this parameter can improve the fit of the model, but in many applications the estimated percentage of contaminants is relatively low. The data, reported later, that are simulated by the Ratcliff diffusion model will not include any contaminants.

2. A complete description of the parameter values is available at the first author's home page.

3. As a matter of fact, the equation is daunting, in the sense that the infinite series does not have an analytical solution, so one must resort to numerical solutions.

4. Equations 6 and 9 hold for both correct and error responses (see, e.g., Laming, 1973, p. 192, footnote 7; Link & Heath, 1975).

5. The close conceptual link between sequential sampling models (e.g., the diffusion model) and signal detection theory has also been a motivating factor in the work of Pew (1969), Emerson (1970), Balakrishnan et al. (2002), and Palmer et al. (2005).

6. Ratcliff and Tuerlinckx (2002) studied model recovery of the Ratcliff diffusion model using 250 and 1,000 observations.

7. When the data are generated by the EZ-diffusion model, it is more efficient to include the error RTs in the calculation of *MRT* and *VRT*. However, this may not be the case when the model is misspecified. In addition, the focus on correct RTs is consistent with current practice in experimental psychology. For these reasons, we choose not to include the error RTs in the computations of *MRT* and *VRT*.

8. This test is available in R (R Development Core Team, 2004) as the function agostino.test() in the *moments* package.

9. This experiment was originally designed to study IQ differences in response speed using the diffusion model. Since the effects of IQ were not statistically reliable, the present analysis collapses over participants with relatively low IQ ($n = 32$) and those with relatively high IQ ($n = 44$).

10. We thank Andrew Heathcote for sending us his R routines for fitting the Ratcliff diffusion model.

APPENDIX
R Code for the EZ-Diffusion Model

This appendix lists the R function (R Development Core Team, 2004) that implements the EZ-diffusion model. As mentioned by Rouder and Lu (2005, p. 603), "R is a freely available, easy-to-install, open-source statistical package based on SPlus. It runs on Windows, Macintosh, and UNIX platforms and can be downloaded from www.R-project.org."

The R function, get.vaTer, takes P_c, *VRT*, and *MRT* as input arguments and returns v, a, and T_{er}:

```
get.vaTer = function(Pc, VRT, MRT, s=0.1)
{
  s2 = s^2
  # The default value for the scaling parameter s equals 0.1
  if (Pc == 0)
    cat("Oops, Pc == 0!\n")
  if (Pc == 0.5)
    cat("Oops, Pc == .5!\n")
  if (Pc == 1)
    cat("Oops, Pc == 1!\n")
  # If Pc equals 0, .5, or 1, the method will not work, and
  # an edge correction is required.
  L = qlogis(Pc)
  # The function "qlogis" calculates the logit.
  x = L*(L*Pc^2 - L*Pc + Pc - .5)/VRT
  v = sign(Pc-.5)*s*x^(1/4)
  # This gives drift rate.
  a = s2*qlogis(Pc)/v
  # This gives boundary separation.
  y = -v*a/s2
  MDT = (a/(2*v)) * (1-exp(y))/(1+exp(y))
  Ter = MRT-MDT
  # This gives nondecision time.
  return(list(v, a, Ter))
}
```

Now consider an EZ-diffusion process for which drift rate $v = 0.1$, boundary separation $a = 0.14$, $T_{er} = 0.300$, and s is set at its arbitrary default value of 0.1. With very many observations, this process will result in $MRT = 0.723$, $VRT = 0.112$, and $P_c = .802$ (these values are rounded). To illustrate and check the above code, the following command may be executed at the R prompt:

```
pars = get.vaTer(.802, .112, .723)
```

Typing "pars" at the R prompt will then display the following:

```
[[1]]
[1] 0.09993853
[[2]]
[1] 0.1399702
[[3]]
[1] 0.30003
```

These values correspond to v, a, and T_{er}, respectively. The code above can of course also be easily implemented in programs such as SPSS or Excel. A JavaScript program that implements the EZ-diffusion model can be found at users.fmg.uva.nl/ewagenmakers/EZ.html.

(Manuscript received October 31, 2005;
revision accepted for publication June 12, 2006.)