Psychonomic Bulletin & Review
2008, 15 (6), 1229-1235
doi:10.3758/PBR.15.6.1229

EZ does it! Extensions of the EZ-diffusion model

Eric-Jan Wagenmakers, Han L. J. van der Maas, Conor V. Dolan, and Raoul P. P. P. Grasman
University of Amsterdam, Amsterdam, The Netherlands

In this rejoinder, we address two of Ratcliff's main concerns with respect to the EZ-diffusion model (Ratcliff, 2008). First, we introduce "robust-EZ," a mixture model approach to achieve robustness against the presence of response contaminants that might otherwise distort parameter estimates. Second, we discuss an extension of the EZ model that allows the estimation of starting point as an additional parameter. Together with recently developed, user-friendly software programs for fitting the full diffusion model (Vandekerckhove & Tuerlinckx, 2007; Voss & Voss, 2007), the development of the EZ model and its extensions is part of a larger effort to make diffusion model analyses accessible to a broader audience, an effort that is long overdue.

In our original article (Wagenmakers, van der Maas, & Grasman, 2007), we extolled the virtues of the Ratcliff diffusion model for speeded two-choice tasks (Ratcliff, 1978; Ratcliff & McKoon, 2008; Wagenmakers, in press). In particular, we stressed that the model allows researchers to decompose observed performance (i.e., both response times [RTs] and proportions correct) into unobserved psychological processes. The most important psychological processes in the diffusion model are speed of information accumulation (i.e., drift rate v), response caution (i.e., boundary separation a), and time spent on processes, such as encoding and motor execution, that are not directly involved in discriminating between response alternatives (i.e., nondecision time T_{er}). We argued that a diffusion model decomposition of performance is statistically more appropriate and theoretically more meaningful than the analyses that are currently standard in the field.

Despite its generality and impressive track record, however, only a few researchers have applied the diffusion model in their own work. We believed that the reason for this state of affairs is twofold. First, at the time of writing, no model-fitting software was publicly available, and this meant that only a select group of researchers—those with the technical skill to program the required routines from scratch—were able to apply the model to data. The need for technical skill becomes apparent when we consider that a *simplified* version of the full Ratcliff diffusion model[1] requires the evaluation of the following integral (Ratcliff & Smith, 2004, Equation A13; Tuerlinckx, 2004) given in Equation 1 below, where $U_{\pm r}$ denotes the uniform distribution on the interval $(-r, r)$, N denotes the normal distribution, and G denotes the probability of an error response before time t, which (as Cox & Miller, 1970, described) is given by Equation 2. Here, P_e is the probability of an error, irrespective of the time when it was committed; ξ, ζ, and τ are, respectively, the drift rate, starting point, and nondecision time on a given trial; and v, z, and T_{er} are, respectively, the across-trials *expected* drift rate, starting point, and nondecision time. Note that Equation 1 contains a triple integral over trial-to-trial variability in nondecision time (T_{er}), a priori bias or starting point (z), and drift rate (v). Furthermore, Equation 2 contains an infinite sum. The expected time and effort associated with the implementation of this model may well have discouraged many researchers from pursuing it.

There is another reason why the diffusion model has not been widely adopted outside of a narrow circle of mathematical psychologists: The model is difficult to apply when the number of observations in each condition is low. One rule of thumb is that each condition should have about 10 error RTs for the model fit to be reliable.[2] Ratcliff (2008) provided two interesting exceptions that prove the rule. It is unfortunate that researchers, convinced of the virtues of the diffusion model, subsequently have to accept that they can apply the model only if their sample sizes are much larger than those they routinely collect. For instance, experiments in visual word recognition often feature 25 or fewer observations per condition per participant; application of the rule of thumb would require participants to have an unacceptably high error rate of 40%.

To bridge the gap between what experimental psychologists need and what the full Ratcliff diffusion model requires, we developed the "EZ"-diffusion model (Wagenmakers et al., 2007). The EZ model takes three

$$\iiint_{-\infty}^{\infty} G(t,\xi,\zeta,a,\tau)\, U_{\pm\frac{s_t}{2}}(\tau - T_{er})\, U_{\pm\frac{s_z}{2}}(\zeta - z)\, N(\xi - v, \eta^2)\, d\tau d\zeta d\xi \tag{1}$$

$$G(t,\xi,\zeta,a,\tau) = P_e - \frac{\pi s^2}{a^2}\exp\left(\frac{-\zeta\xi}{s^2}\right)\sum_{k=1}^{\infty}\frac{2k\sin\left(\frac{\pi k\zeta}{a}\right)\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)(t-\tau)\right]}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)} \tag{2}$$

E.-J. Wagenmakers, ej.wagenmakers@gmail.com

observed quantities from each condition—namely, the mean RT for correct responses (i.e., *MRT*), the variance in RTs for correct responses (i.e., *VRT*), and the proportion correct (i.e., P_c)—and transforms these quantities into estimates for drift rate v, boundary separation a, and nondecision time T_{er}. The one-to-one transformations are achieved using closed-form equations and do not require an iterative fitting program.[3] The tractability of the transformations does come at a cost, however; in its calculations, the EZ model makes two simplifying assumptions—namely, the absence of trial-to-trial variability in v, z, and T_{er}, and the absence of any a priori response bias toward one of the two choice alternatives (i.e., starting point z is equal to $a/2$).

We hoped that the EZ-diffusion model would help popularize the full diffusion model; when researchers use the EZ model and experience the advantages of a diffusion model analysis firsthand, they might then start to explore methods that are more sophisticated but that are also more demanding, in the sense of requiring more data and a higher level of statistical expertise. We emphasize, as we did in our original article, that our aim was not to offer a substitute for the full Ratcliff diffusion model; rather, our aim was to provide a means to obtain a rough-and-ready estimate of the underlying psychological processes, an estimate that could, if possible, later be improved upon through a more complete analysis.

In a response published in this journal, Ratcliff (2008) criticized our original EZ article and the EZ model on several counts. In this rejoinder, we ignore some of the more philosophical differences of opinion between Dr. Ratcliff and ourselves. Instead, we wish to focus on what we perceive to be the most practically relevant concerns—namely, the imperfections and limitations of the EZ method that were highlighted in the Ratcliff simulations. In our original EZ article, we had already identified some of these imperfections and limitations; nevertheless, the Ratcliff (2008) article contains new results that warrant a closer examination.

The Ratcliff Simulations

The simulations reported in Ratcliff (2008) cover a lot of ground, but they also share a common theme. This theme is to take the full Ratcliff diffusion model, to generate data using components that the EZ model had discarded to achieve tractability, and then to show that the EZ model poorly recovers the parameters that generated the data. Thus, the EZ model assumes the absence of "response contaminants" (Ratcliff & Tuerlinckx, 2002). Ratcliff's simulations showed that when the data do have response contaminants, the EZ model may perform poorly. The EZ model also assumes that the starting point is exactly in-between the response boundaries, so that there is no a priori bias. Ratcliff (2008) showed that when there is a priori bias, the EZ model may again perform poorly. Finally, the EZ model assumes the absence of trial-to-trial variability in drift rate, starting point, and nondecision time. Ratcliff (2008) showed that when such variability is present, the EZ model may, once again, perform poorly. None of this is surprising, and in fact, much of it was anticipated and acknowledged in the original EZ article (Wagenmakers et al., 2007). In general, it is easy to generate data from a complex model and show that the simpler model, nested within the complex model, fails to recover parameters well.

Nevertheless, we concede that the demonstrated sensitivity of the EZ model to the presence of contaminants is worrying, and that the assumption that the starting point is exactly between the two response boundaries may be overly restrictive. Consequently, we have extended the EZ model to address both issues, producing the "robust-EZ" model.

EZ extension 1: Robustness against contaminants. Ratcliff and Tuerlinckx (2002) put forward the notion that, on a certain proportion of trials, participants may experience a temporary lapse of attention. Processing of the stimulus stops when the attentional lapse starts, and resumes without loss of information when the attentional lapse ends. This means that lapses do not influence the proportion of errors; lapses, or *response contaminants*, only add a variable delay to the RT that would otherwise have been observed, irrespective of whether the RT ends up belonging to a correct or an error trial.

Ratcliff (2008) showed that the presence of response contaminants can greatly bias the parameter estimates of the EZ model. To increase the robustness of the EZ model to the presence of response contaminants, we have followed Ratcliff and Tuerlinckx (2002) in proposing a mixture modeling approach.

We assume that the observed distribution of RTs is a mixture of two components. The first represents the response contaminants and follows a uniform distribution that ranges from the fastest to the slowest RT. The second component represents the RTs of interest—namely, those that are generated by the diffusion model. We assume that this second component is adequately captured by the ex-Gaussian distribution (see, e.g., Ratcliff, 1978, 1979).

As the name suggests, the ex-Gaussian distribution is the additive combination of an exponentially distributed random variable with rate $1/\tau$ and a Gaussian, or normally distributed, random variable with mean μ and standard deviation σ (Hohle, 1965; Luce, 1986). Although the ex-Gaussian distribution may not be very satisfactory from a theoretical perspective, it is relatively easy to use, and it generally produces a good fit to empirical RT distributions (see, e.g., Ratcliff, 1979, Figure 3).

Thus, the mixture model involves the estimation of four parameters—that is, the mixture proportion α and the ex-Gaussian parameters μ, σ, and τ. From the last three parameters, one can calculate the mean and the variance of the ex-Gaussian distribution. The idea of the robust-EZ model is to use as input to the EZ equations not the *observed* sample moments *MRT* and *VRT*, but the *inferred* mean and variance from the ex-Gaussian component of the mixture model—that is, MRT_{EG} and VRT_{EG}. In this way, the mixture model filters out the effects of contaminants on the observed sample *MRT* and *VRT*, and thereby greatly reduces the impact of response contaminants on the parameter estimates. The

Appendix describes in more detail both the fitting procedure and the associated software in which the model complexities are hidden from the user. It is true that robust-EZ is slightly more involved than the original EZ model. Still, the only thing users have to do to obtain robust-EZ parameter values is to construct a file that lists the total number of observations and the correct RTs, and then to call the fitting routine by issuing a single command.

In order to ascertain that robust-EZ works as intended, we compared its performance against that of the original EZ model for eight synthetic data sets of 10,000 observations each (see Ratcliff, 2008, for a similar procedure). For data generation, boundary separation a was fixed at 0.12, and nondecision time T_{er} was fixed at 0.300 sec. Four data sets had a relatively low value for drift rate (i.e., $v = 0.1$), and the other four data sets had a high value for drift rate (i.e., $v = 0.3$). Each of the two drift rates was associated with four levels of contamination: 0%, 2%, 5%, and 10%. Specifically, in each data set, x% of EZ-generated correct RTs were "contaminated" by adding an RT that was drawn from a uniform distribution ranging from 0 to 2 sec. Finally, RTs higher than 2.5 sec were eliminated from the analyses.

Figure 1 shows the effects of contamination both for the original EZ model and for robust-EZ. In agreement with Ratcliff, it is clear that the original EZ model is sensitive to the presence of contaminants. Specifically, contaminants lead to an overestimation of boundary separation (i.e., middle panels) and underestimations of drift rate (i.e., left panels) and nondecision time (i.e., right panels). This pattern is more pronounced when drift rate is high than when it is low. Robust-EZ, as one would expect, is not much affected by the insertion of contaminants.

The foregoing may be viewed as a *proof of principle*, in that the simulation shows that the robust-EZ procedure is able—for very large data sets—to correctly estimate the EZ parameters, even in the presence of contaminants. Of course, practically relevant settings feature many fewer observations, and in this case performance of the robust-EZ procedure will decrease because of sampling error. We conducted two additional simulations to study the extent to which robust-EZ can recover parameter values when the number of observations is low.

In the first additional simulation, we generated data from a single-condition EZ model with these parameters: drift

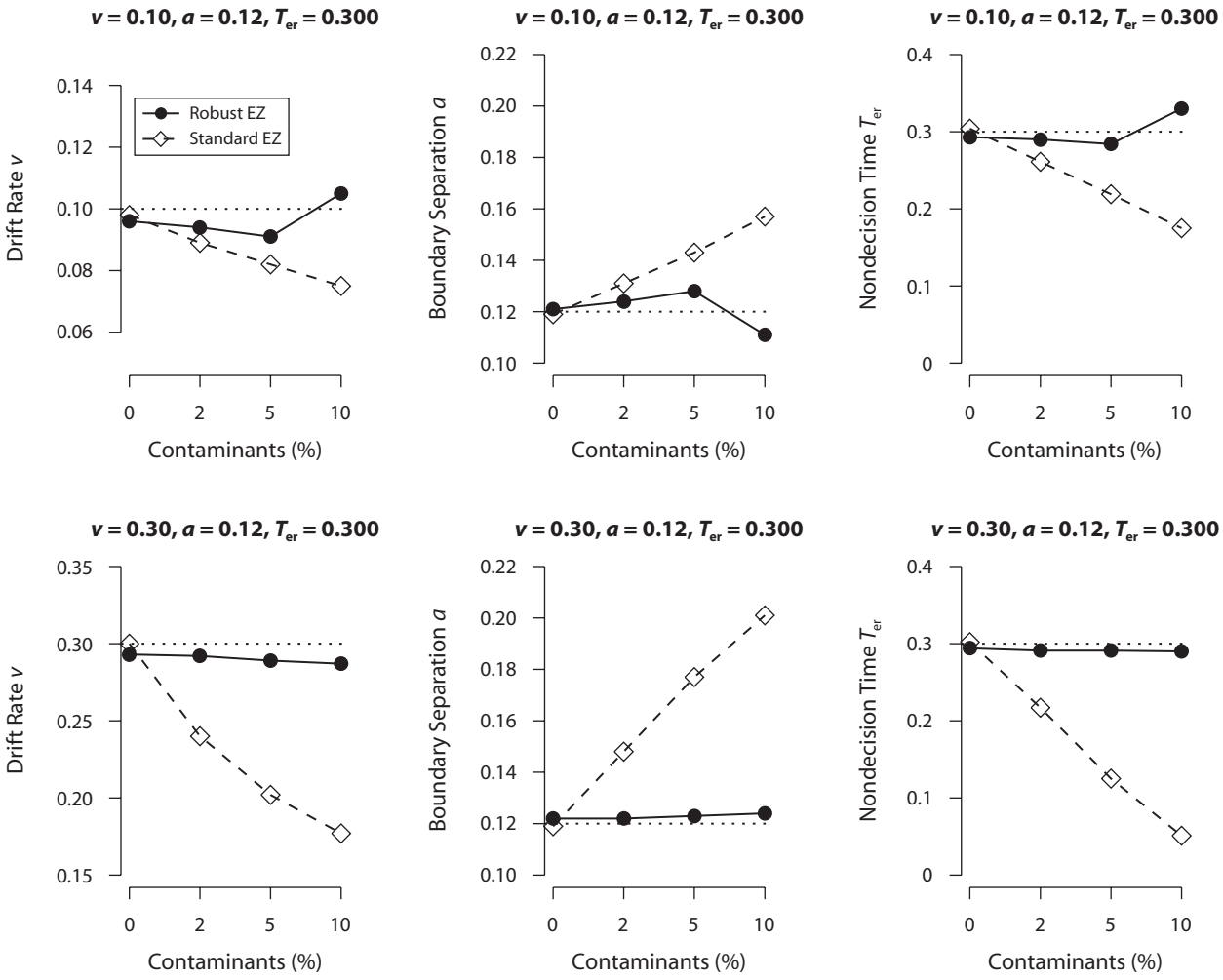

Figure 1. Parameter estimation of the standard EZ method versus the robust-EZ method under different levels of contamination. The horizontal dotted lines indicate the true parameter values.

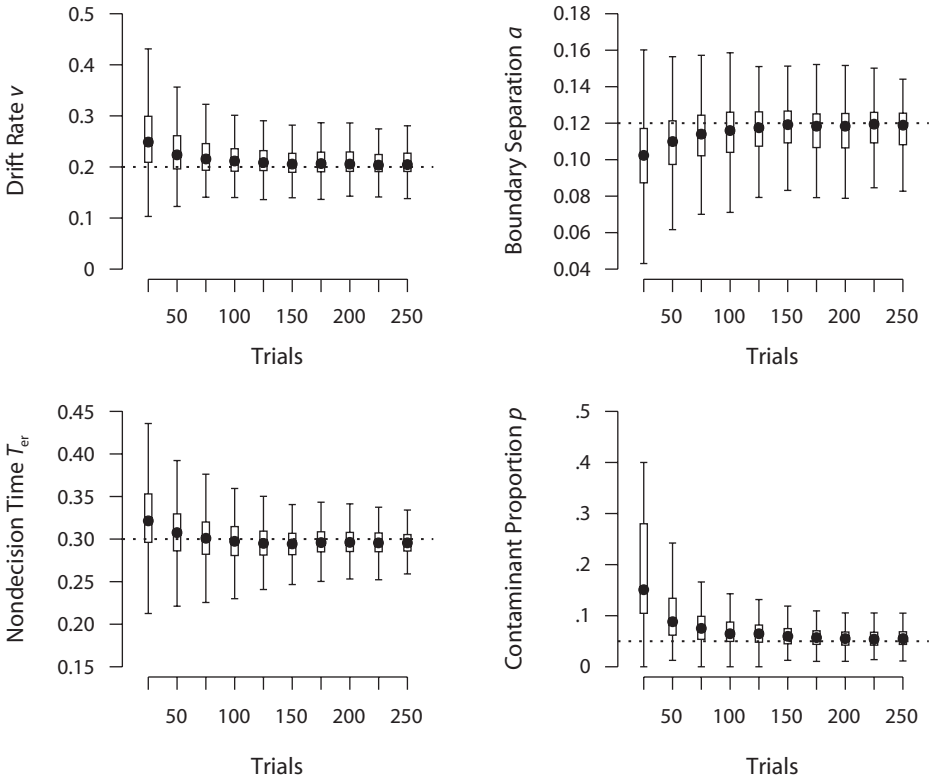

Figure 2. Parameter estimation of the robust-EZ procedure for a single participant in a single condition, as a function of the total number of trials. The horizontal dotted lines indicate the true parameter values.

rate $v = 0.2$, boundary separation $a = 0.12$, nondecision time $T_{er} = 0.300$, and the proportion of contaminants $p = .05$ (uniformly distributed between 0 and 2 sec). We conducted 1,000 synthetic experiments, for each of 10 different numbers of trials, ranging from $N = 25$ to $N = 250$ in steps of 25. For this single-participant scenario, the parameter recovery is shown in Figure 2, in which each panel plots 10 box-and-whiskers plots, 1 for each N, N, ranging from 25 to 250 in steps of 25. A box-and-whiskers 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 quartile to the .75 quartile, and the dot in the middle of the box is the .50 quartile (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 in Figure 2, for $N \geq 100$, parameter recovery is generally unbiased; that is, the median of the recovered parameter values (i.e., the dots in the boxes) is close to the horizontal line that indicates the parameter value with which the data were generated. For $N < 100$, the bottom right panel shows that the proportion of contaminants is overestimated, and this is associated with an overestimation of drift rate and nondecision time, as well as an underestimation of boundary separation.

Figure 2 also shows that, in the case of low N, parameter recovery is relatively imprecise. This problem becomes less serious as N grows large, but even for $N = 250$, parameter estimation comes with nonnegligible uncertainty. However, the goal of experimental work is often not to estimate parameters for a single participant; rather, the goal is almost always to draw conclusions from a group average of individual parameter estimates. To highlight the extent to which parameter uncertainty can be reduced by considering more than a single participant, a second simulation replicated the first, but now we averaged parameters over 30 synthetic participants.

In this second additional simulation, we assumed the following three independent sources of between-participants variability: Drift rate v was uniformly distributed between 0.15 and 0.25, boundary separation a was uniformly distributed between 0.10 and 0.14, and nondecision time T_{er} was uniformly distributed between 0.250 and 0.350. For all synthetic participants, the proportion of contaminants was fixed at .05 (uniformly distributed between 0 and 2 sec). We again conducted 1,000 experiments, for each of 30 participants and for each of 10 different numbers of trials, ranging from $N = 25$ to $N = 250$ in steps of 25.

Figure 3 shows the results. As is evident from a comparison of Figures 2 and 3, the uncertainty in estimation is much smaller when parameters are averaged across several participants, despite the presence of participant-to-participant variability in all three EZ parameters. Figures 2 and 3 both show the same qualitative pattern of bias for $N < 100$: an overestimation of the proportion of contaminants, with an associated overestimation of drift rate and nondecision time, and an underestimation of boundary

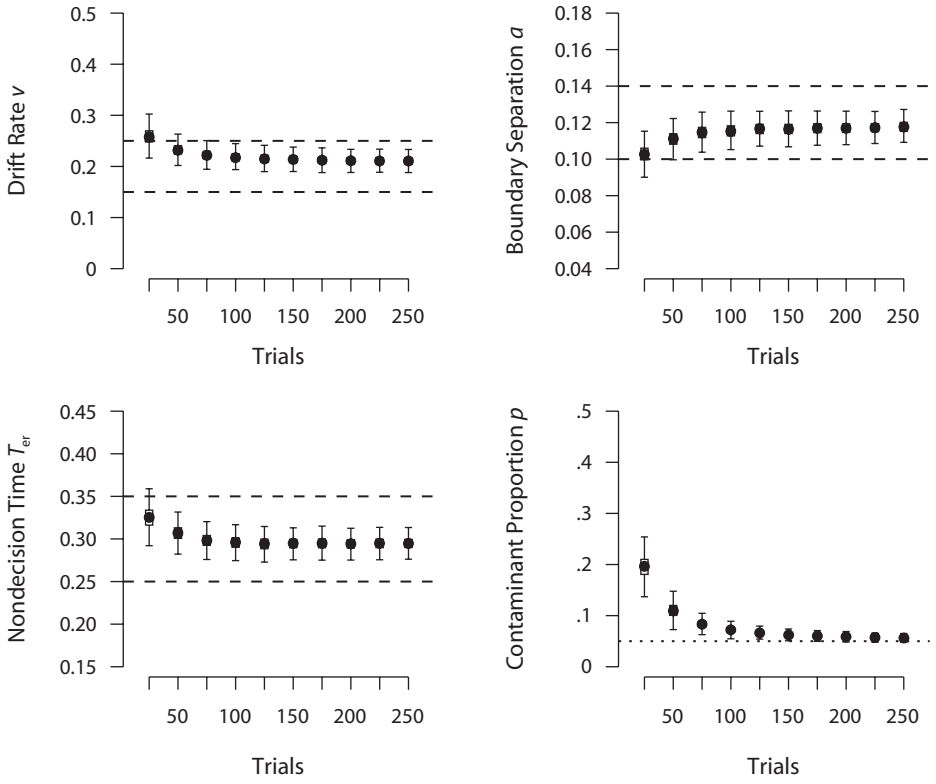

Figure 3. Average parameter estimation of the robust-EZ procedure for 30 participants in a single condition, as a function of the total number of trials. In the panels for drift rate, boundary separation, and nondecision time, the two horizontal dashed lines indicate the upper and lower bounds for the uniform distribution that describes between-participant variability. In the panel for the proportion of contaminants, the horizontal dotted line indicates the true parameter value. Each box plot is based on 1,000 replications.

separation. When $N \geq 100$, parameter estimation is relatively unbiased and precise. Increasing the number of trials beyond 100 does not appear to yield much improvement.

The previous two simulations provide a rough indication of the amount of variability that might be expected when the robust-EZ procedure is applied to small-N data sets such as those that often occur in experimental psychology. Of course, the results shown in Figures 2 and 3 hold for a particular, albeit representative, set of generative parameters. Also note that with few observations per participant, few participants, and large intraparticipant variability, one can only hope to detect very large experimental effects; this is as true for robust-EZ as for any other statistical model.

EZ extension 2: Estimation of response bias. Grasman, Wagenmakers, and van der Maas (in press) recently proposed an extension of the EZ model that allows the starting point to vary freely and parameters to be constrained across conditions.[4] The key equation of the extended model gives the variance of the decision time distribution for a diffusion process with a priori bias, conditional on the response boundary that was reached first. A software program that implements this EZ extension is freely available online at the Web address provided in note 4.

The extension to starting point estimation was made possible by the fact that in a speeded two-choice task, there are two stimulus categories that may have different drift rates yet share values for decision criteria. Thus, for example, for old and new words in a recognition memory task, the extended model takes as input P_c^{old}, P_c^{new}, VRT^{old}, VRT^{new}, MRT^{old}, and MRT^{new} and returns as output estimates for v^{old}, v^{new}, a, z, T_{er}^{old}, and T_{er}^{new}. Note that a and z are response criteria that are assumed to be determined prior to stimulus processing, and so are independent of whether the stimulus word is old or new. If desired, the online program can determine a common estimate for T_{er} using a least-squares fitting procedure.

Other Recent Developments in Fitting the Diffusion Model

Since the introduction of EZ a year ago, two computer programs that greatly facilitate diffusion model analyses have been made publicly available. These programs are DMAT,[5] a MATLAB toolbox by Joachim Vandekerckhove and Francis Tuerlinckx (2007, 2008), and fast-dm,[6] a platform-independent command line tool developed by Andreas and Jochen Voss (2007, 2008). DMAT and fast-dm both use the original EZ parameters as default starting values for their optimization routines. These excellent programs are well-documented, are user-friendly, and come free of charge.

Although these programs should be used whenever possible—something that Dr. Ratcliff fully agrees with—

many situations remain in which the EZ diffusion model and its extensions might be useful. For instance, DMAT ignores conditions with fewer than 11 errors; both DMAT and fast-dm require more effort on the part of the user than does EZ, both in terms of learning to operate the program and learning to interpret the results statistically; and DMAT and fast-dm can be slow to fit many conditions, whereas EZ estimates are instantaneous. In our own work, we have used EZ estimates to provide online feedback with respect to drift rate and boundary separation, as well as to provide immediate estimates of nondecision time to better adjust speed–accuracy requirements in a comparison of young and old adults. In addition, EZ estimates can be useful as a preliminary check to see whether the data are interesting enough to model in a more detailed fashion. For instance, an EZ analysis of data from an fMRI experiment (Forstmann et al., in press) showed that a manipulation of speed–accuracy instructions was associated with changes in boundary separation. Moreover, the EZ analysis showed that people who had a relatively large decrease in boundary separation under speed stress also had a relatively large increase in activation of the anterior striatum and the pre-supplementary motor area, a result that is in line with neurocomputational models (e.g., Lo & Wang, 2006). A more detailed analysis with the linear ballistic accumulator model (Brown & Heathcote, in press) was consistent with these initial EZ findings.

Conclusions

The simulations reported by Ratcliff (2008) demonstrate that when the simplifying conditions of the EZ model are violated, it is overly sensitive to the presence of contaminants, is biased in the face of misspecification, and is less efficient then the chi-square method (Ratcliff & Tuerlinckx, 2002). We agree with Ratcliff (2008) that the sensitivity to contaminants is a potential pitfall, and we have introduced the robust-EZ method to address the issue. Misspecification and lower efficiency in estimation are unavoidable consequences of simplicity; nevertheless, we have recently developed an extension of the EZ model that allows starting point to vary freely between conditions (Grasman et al., in press).

We believe that experimental psychology can greatly profit from widespread adoption of the diffusion model as an analytical tool. For this to happen, the existence of user-friendly fitting routines is absolutely essential. We see the development of methods such as the EZ-diffusion model as a valuable first step toward making the diffusion more accessible to a broader audience of researchers. From this perspective, the EZ model accomplishes its goals.

AUTHOR NOTE

Correspondence concerning this article may be addressed to E.-J. Wagenmakers, University of Amsterdam, Department of Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands (e-mail: ej.wagenmakers@gmail.com).

REFERENCES

BROWN, S. D., & HEATHCOTE, A. J. (in press). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*.

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

DOLAN, C. V., VAN DER MAAS, H. L. J., & MOLENAAR, P. C. M. (2002). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods, Instruments, & Computers*, **34**, 304-323.

FORSTMANN, B. U., DUTILH, G., BROWN, S., NEUMANN, J., VON CRAMON, D. Y., RIDDERINKHOF, K. R., & WAGENMAKERS, E.-J. (in press). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*.

GRASMAN, R. P. P. P., WAGENMAKERS, E.-J., & VAN DER MAAS, H. L. J. (in press). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology*.

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

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

LO, C.-C., & WANG, X.-J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience*, **9**, 956-963.

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

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. (2008). The EZ diffusion method: Too EZ? *Psychonomic Bulletin & Review*, **15**, 1218-1228.

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

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

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.

R DEVELOPMENT CORE TEAM (2004). *R: A language and environment for statistical computing.* Vienna: Author.

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). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

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

VANDEKERCKHOVE, J., & TUERLINCKX, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, **40**, 61-72.

VOSS, A., & VOSS, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, **39**, 767-775.

VOSS, A., & VOSS, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology*, **52**, 1-9.

WAGENMAKERS, E.-J. (in press). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*.

WAGENMAKERS, E.-J., VAN DER MAAS, H. J. L., & GRASMAN, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, **14**, 3-22.

NOTES

1. The complete version is a mixture of a diffusion model component and a component intended to capture response contaminants.

2. This particular rule of thumb is made explicit in the current implementation of the DMAT software program, discussed later, which ignores error RTs when they are fewer than 11 in number. Ratcliff and Tuerlinckx (2002) mention 5 as an absolute minimum number of error RTs required to fit the model at all.

3. The Web site users.fmg.uva.nl/ewagenmakers/papers.html lists EZ implementations in Excel, R, and JavaScript.

4. The manuscript (currently accepted pending minor revisions for the *Journal of Mathematical Psychology*), R code, and JavaScript code are available at purl.oclc.org/net/rgrasman/ez2.

5. DMAT is freely available online at ppw.kuleuven.be/okp/dmatoolbox.

6. Fast-dm is freely available online at www.psychologie.uni-freiburg.de/Members/voss/fast-dm.

APPENDIX
The Robust-EZ Model

Robust-EZ fits to data a mixture model of two components: a uniform distribution of response contaminants, ranging from the fastest to the slowest RT, and an ex-Gaussian distribution that represents the process of interest. The ex-Gaussian distribution is the additive combination of an exponential distribution with rate $1/\tau$ and a Gaussian or normal distribution with mean μ and standard deviation σ.

As starting values for the parameter optimization routine, robust-EZ uses the method-of-moment estimators for the ex-Gaussian—that is,

$$\mu = M_1 - \tau, \tag{A1}$$

$$\sigma = \sqrt{M_2 - \tau^2}, \tag{A2}$$

and

$$\tau = \left(\frac{1}{2} M_3\right)^{\frac{1}{3}}, \tag{A3}$$

in which M_1, M_2, and M_3 denote mean, variance, and skew, respectively (Heathcote, 1996).

Based on *disfit* routines developed by Dolan, van der Maas, and Molenaar (2002), robust-EZ was implemented in the R environment[A1] to fit the uniform/ex-Gaussian mixture with the method of maximum likelihood, using Equations A1–A3 for starting values and a quasi-Newton method for optimization.

After completing the optimization process, robust-EZ uses the maximum likelihood estimates $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\tau}$ to calculate the estimated mean and variance of the component process of interest, as $MRT_{EG} = \hat{\mu} + \hat{\tau}$ and $VRT_{EG} = \hat{\sigma}^2 + \hat{\tau}^2$. Finally, robust-EZ uses MRT_{EG}, VRT_{EG}, and P_c to determine drift rate, boundary separation, and nondecision time, as per Equations 5–9 in Wagenmakers et al. (2007).

Clearly, this procedure is more involved than the original EZ-diffusion model. However, we have implemented robust-EZ in such a way that the model complexity is hidden from the user. The process works as follows. First, the user constructs a file that, on the first line, lists the number of trials from a specific experimental condition (e.g., 200). Each of the lines below contains a single RT (in seconds) for correct responses only.

In order to illustrate the process, we have constructed the file ExampleDataREZ.txt, which is based on 200 observations from an EZ-diffusion process with drift rate $v = 0.2$, boundary separation $a = 0.12$, and nondecision time $T_{er} = 0.300$. These are the first five lines of ExampleDataREZ.txt:

```
200
0.56345
0.53415
0.39205
0.64095
```

Once the user has formatted the data in this way, the only thing left to do is to type, on the R command line, the RobustEZ.from.File function calling the file with the data:

```
> RobustEZ.from.File("ExampleDataREZ.txt")
```

The final line of the output then contains four numbers:

```
[1]  0.2139106  0.1174465  0.3028974  0.9685783
```

These four numbers (left to right) are the robust-EZ estimates for drift rate, boundary separation, nondecision time, and mixture proportion. Here, the data were generated without contaminants, so the true value of the mixture proportion is 1. The files that come with the robust-EZ software are available on the first author's Web page.

NOTE

A1. R (R Development Core Team, 2004) is a platform-independent software package for statistical computing, and it can be downloaded free from www.r-project.org.

(Manuscript received April 29, 2008;
revision accepted for publication August 1, 2008.)