



The EZ Diffusion Model: An overview with derivation, software, and an application to the Same-Different task 🗨️

Julien T. Groulx ^a , Bradley Harding ^b & Denis Cousineau ^c

^aUniversité du Québec à Montréal

^bUniversité de Moncton

^bUniversité d'Ottawa

Abstract ■ The diffusion model is useful for analyzing data from decision making experiments as it gives information about a dataset that regular statistical tests cannot, including: the rate of processing, the encoding and motor response times, and decision thresholds. The EZ diffusion model is a restricted version of the diffusion model with some parameter variability set to zero, allowing for quicker analyses. Here we describe the EZ diffusion model –including how it was derived mathematically– the measurement units of the parameters, and how it can be generalized to starting points other than the mid-point. We also show how its parameters can be estimated using computer software (the model is available with many software programs such as R and Excel, to which we add SPSS and a Mathematica code). Finally, an EZ analysis was run on one dataset obtained from a “Same”-“Different” experiment.

Keywords ■ Diffusion model; EZ; response times. **Tools** ■ SPSS, Mathematica.

denis.cousineau@uottawa.ca

[10.20982/tqmp.16.2.p154](https://doi.org/10.20982/tqmp.16.2.p154)

Acting Editor ■
[Cheng-Ta Yang](#)
(Department of Psychology, National Cheng Kung University)

Reviewers
■ Two anonymous reviewers

Introduction

Cognitive psychology tries to explain the brain’s fundamental capabilities by studying its various mental processes, acknowledging the dynamic interplay of these processes, and attempting to model them stochastically as best as possible. For example, such models can be used for a better understanding of theoretical mechanisms related to memory, decision-making, and learning. These models, based on a pre-established theory, are tested on samples of data. If the model fits consistently over multiple studies, the theory may come to be accepted as a reliable explanation. Regarding decision-making paradigms, models are often based on accumulation of evidence or diffusion models, or as they are more generally called, *sampling models* (Smith & Ratcliff, 2015). These models represent decisions as the accumulation of evidence until a threshold is reached, leading to the corresponding action (Luce, 1986; Townsend & Ashby, 1983).

Response time (RT) and accuracy (ACC) are the two

measured variables in simple decision-making experiments. These data are typically analyzed with statistical tools such as ANOVA. However, Wagenmakers, Van Der Maas, and Grasman (2007) argue that sampling models allow interpretation usually impossible with ANOVA or similar methods. Neither RTs nor ACCs alone can represent the internal dynamics underlying a subject’s performance; the interrelation between these two variables must be considered. For example, a subject with fast RTs and low ACCs does not necessarily have a better overall performance than a subject with slow RTs and high ACCs. Sampling models are useful to assess these discrepancies as they allow, by analyzing performance holistically, the determination of latent variables, which can improve the modeling of performance.

Donkin, Brown, Heathcote, and Wagenmakers (2011) support that “A growing number of researchers use cognitive models of choice response time (RT) to draw conclusions about the psychological processes that drive decision making.” (p. 61). According to them, an expla-



nation must provide information about three key aspects of behavior: response bias, stimulus saliency and non-decision time. These concepts are represented by unobservable variables: response caution (distance between starting point z and response boundaries a), mean rate of evidence accumulation (or, drift rate; v), and time related to non-decision processes (T_{er}).

There are multiple models used in cognitive science. However, this manuscript will focus solely on the EZ diffusion model and present its general properties, its usefulness, and possible liability. To do so, we first present a comparison with the more general model, the Ratcliff's diffusion model (RDM; sometimes called the drift-diffusion model; Ratcliff, 1978). Second, we review the EZ diffusion model, describe its derivation, and indicate the measurement units of the parameters. Furthermore, in the EZ framework, it is not obligatory that the starting point be halfway between the two boundaries; thus, we briefly generalize EZ to other starting positions. Third, we discuss packages and computer code that perform EZ analyses. Two have been previously published (EZ2, a package for R, and an Excel spreadsheet –both available on Raoul Grasman's web site dedicated to EZ: <http://raoul.socsci.uva.nl/EZ2/>) and two additional alternatives are provided with this article (EZ4Math, a package for Mathematica, and a syntax for SPSS). Fourth, we describe an algorithm that simulates a diffusion process akin to the one fitted by EZ. This approach is slower but can be useful for other EZ-related endeavours, such as making figures. Fifth, we provide a measure of fit for EZ so that it can be compared to other models. Finally, we test the EZ diffusion model in a "Same"- "Different" experiment, a task in which participants are required to answer as rapidly and as accurately as possible whether two successive stimuli are the "Same" or "Different"; some preliminary results are presented.

Ratcliff's Diffusion Model

We present the RDM first as EZ is a restricted version of this general model; all components originate from the former. The RDM was first introduced by Ratcliff in 1978 and has shown useful for describing performance that RTs and ACCs cannot divulgate by themselves (Ratcliff, 1978; Wagenmakers et al., 2007). The model has been tested and generally fits well to a variety of data stemming from many different experiments such as lexical decision tasks, recognition tasks, etc. It is sometimes referred to as a type of *Random Walk model*, although the latter is based on discrete steps whereas the diffusion model is based on a continuous accumulation of evidence. The model has two boundaries, each triggering a certain response (e.g., the "Yes" and the "No" responses). In a "Same"- "Different" experiment, one boundary would trigger the "Same" re-

sponse and the other the "Different" response. The evidence accumulation is a process that fluctuates from a starting point towards one of two boundaries. When the evidence count reaches a boundary, the corresponding decision is made for that specific trial. A central assumption of the diffusion model is that within each trial, the accumulation rate varies according to a distribution with a mean drift rate and a certain variability about this mean drift rate given by the stimulus' characteristics. If within-trial variability is null, the evidence accumulation would rise linearly with time. In Figure 1, we illustrate a few trials with within-trial variability according to a diffusion model.

To fit the diffusion model to a set of data, this model requires the decision made on all trials. If accuracy is available, it must be transformed so that "top" responses are coded with "1" and "bottom" responses are coded with "0". It also requires the raw RT for all trials (either in seconds or milliseconds –see later for a note on measurement units) and analyses must be performed on a per-subject basis. However, this may cause issues for certain paradigms as per-subject analyses require a large amount of trials to get reliable estimates (in the order of hundreds of trials; see Lerche & Voss, 2018, for recommendations on sample size). This is also problematic if incorrect answers are rare (or more precisely, if one boundary is much less frequently reached than the other). Wagenmakers et al. (2007) noted:

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. (p. 6)

With this constraint, a typical experiment can hardly have more than 4 conditions within a session of 60 minutes. Testing would be unrealistically long if more conditions were added.

The diffusion model parameters can be estimated with, for example, *fast-dm*, a standalone program (Voss & Voss, 2007) or *DMAT*, a toolbox for *Matlab* (Vandekerckhove & Tuerlinckx, 2008). These software packages were evaluated in van Ravenzwaaij and Oberauer (2009). However, it can take many hours to get estimates from the data of a single participant as the whole estimation process is computationally demanding (but see Navarro & Fuss, 2009; Tuerlinckx, 2004). More recently, the hierarchical diffusion model (HDM) and the hierarchical drift diffusion model (HDDM) were developed to fit a diffusion model within a hierarchical framework (Vandekerckhove, Tuerlinckx, & Lee, 2011; Wiecki, Sofer, & Frank, 2013, respectively).

The RDM requires 8 parameters, as listed in Table 1. The first parameter is the *mean drift rate* (v). This param-

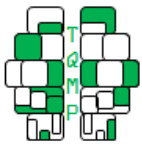
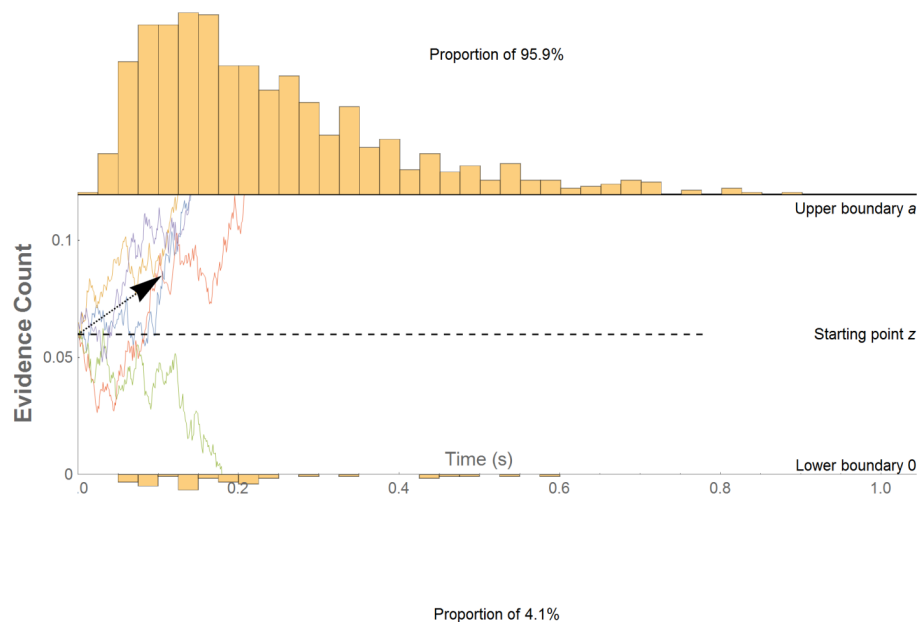


Figure 1 ■ Center: five trials in which the evidence accumulation process is visible as a function of time. Four of the trials ended with a "Top" response. The simulated trajectories are based on George's parameters, a fictitious character introduced in Wagenmakers, Van Der Maas, and Grasman (2007), whose parameters are $\nu = 0.2513e/s$, $a = 0.1197e$, $T_{er} = 0.3012s$, assuming that $s = 0.1e/\sqrt{s}$. Top: distribution of decision times over 1 000 trials, 95.9% of which ended by crossing the top barrier; Bottom: distribution of decision times for the remaining 4.1% of the trials which finished by crossing the bottom barrier.



eter represents the mean rate at which evidence (information leading to either of the boundaries) is accumulated. As can be observed by Figure 1, in the diffusion model, obtaining evidence towards one decision always implies going away from the other decision. The value of this parameter depends on the stimulus presented: A stimulus with high saliency should presumably result in a higher mean drift rate. Positive drift rates will most likely reach the upper boundary whereas negative drift rates will tend to reach the lower boundary. The mean drift rate can be seen as an arrow having a certain slope pointing towards one of the two boundaries (Ratcliff & Smith, 2015). However, the actual path taken is altered by random fluctuations as shown in Figure 1, representing the actual accumulation of evidence towards a decision for five simulated trials.

The second parameter is *across-trial variability in drift rate* (η). This parameter accounts for the fact that different trials might have different time courses and explains why the error RT distribution considerably overlaps the correct RT distribution. The third parameter is the *within-trial variability in drift rate* (s). Without this parameter, the accumulation of evidence would go at a constant rate, with

no random fluctuations. When this is the case, the model is called a *Linear Ballistic model* (such as the LBA; introduced in Brown & Heathcote, 2008). Whereas η changes the rate of evidence accumulation across trials, s changes the amount of evidence accumulated at every moment within a trial. This last parameter cannot be estimated at the same times as the other parameters because the model's temporal unit is arbitrary. Thus, one parameter linking evidence and time must be set in terms of unit of time so that the other parameters are estimated in the same unit of time. It is conventional to set the s parameter to $0.1e/\sqrt{s}$, where e is the unit describing the amount of evidence (an arbitrary unit) and s stands for second.

The fourth parameter is the *boundary separation* (a). This value also represents the location of the upper boundary because by convention the lower boundary is located at 0. For example, in a brightness discrimination task, the top boundary could trigger the "Light" decision and the bottom one could trigger the "Dark" decision (Ratcliff & Smith, 2015). In Ratcliff's diffusion model, the starting point is not fixed, it can vary from trial to trial depending on the fifth and sixth parameters which are the

**Table 1** ■ Parameters from the diffusion model and how they are included into EZ

Parameter	Parameter meaning	Presence in EZ
ν	Mean Drift Rate	Free
η	Across-Trial Variability in Drift Rate	Fixed to 0
s	Within-Trial variability in drift rate	Conventionally set to $0.1 e/\sqrt{s}$
a	Boundary Separation	Free
z	Mean Starting Point	Fixed to $a/2$
s_z	Across-Trial Variability in Starting Point	Fixed to 0
T_{er}	Mean time of the Non Decision Components	Free
s_t	Across-Trials Variability in the Non-Decision Components	Fixed to 0

mean starting point (z), or the average position on the vertical axis at which the accumulation of evidence begins, and the *across-trial range of variability in starting point* (s_z) respectively. On a given trial, the starting point is selected randomly from a uniform distribution over the range $[z - s_z/2, z + s_z/2]$. These last two parameters allow modeling concepts such as biased decision making when, for example, the subject is instructed to favor one decision rather than the other, which is achieved with asymmetric boundary locations relative to the starting point. It can also model fast errors when variability in starting point brought the starting position very close to the wrong threshold.

The seventh parameter is the mean processing time of the non-decision components (T_{er} stands for *encoding and motor response times*). It is composed of two parts. (i) The times taken for the information to be transported to the decision process (visual information emitted from the screen and entering the eye to be transformed and treated by visual areas, etc.) are collectively called *encoding time* (T_e). (ii) The time once the decision process is completed up to the measuring instrument (motor functions, mechanical signal, etc.) is called *motor response time* (T_r). The non-decision components time is the sum of both these factors: $T_{er} = T_e + T_r$. These non-decision times are added to each decision time. This parameter accounts for the portion of an observed RT which is due to non-decisional processes. The eighth and last parameter captures the *across-trials variability in the non-decision component of processing* ($s_{T_{er}}$, or in short, s_t). It is assumed that across trials, the non-decision times are uniformly distributed with average T_{er} and range $[T_{er} - s_t/2, T_{er} + s_t/2]$.

EZ Diffusion Model

EZ, a simplified version of the RDM, was developed by Wagenmakers et al. (2007; also see Wagenmakers, van der Maas, Dolan, and Grasman, 2008) to (i) promote the use of diffusion models in psychology, and (ii) avoid the need for a strong background in quantitative modeling or the use of

specialized software. It has fewer parameters, and consequently requires less information to fit empirical data than its more complex counterpart. Likewise, the authors report that many researchers do not use the diffusion model even where it is effective because it is too difficult to fit (Wagenmakers et al., 2007; also see Röhner & Thoss, 2018). Consequently, the EZ model was created to have a simplified diffusion model which still captures the crucial variables while being significantly simpler and faster to apply than the full model.

As said above, another problem regarding Ratcliff's model is the need for a large number of raw data per condition and a paradigm where accuracy is not too high. Indeed, smaller datasets lead to more variable parameter estimates and thus to less accurate estimates (Cousineau, 2009; Lerche & Voss, 2018). EZ, on the other hand, has stable estimates even with smaller sample sizes, owing to the fact that it is based on fewer parameters. EZ, like any sampling model, should be fit on a per-participant basis as group characteristics are not necessarily the average of the individuals' characteristics (Cousineau, Harding, Thivierge, & Lacouture, 2015).

As previously mentioned, EZ focuses on the three parameters that are the most pertinent to the decision process, namely: the drift rate (ν), the boundary separation (a) and the non-decision time (T_{er}), taken from Ratcliff's full diffusion model. To do so, EZ fixes the starting point parameter's value z to $a/2$, that is, halfway between the two boundaries. Likewise, the three between-trial variability parameters (η , s_z and s_t) are also removed (i.e., set to 0), because, according to Wagenmakers et al. (2007), these factors have a rather small influence on the results and are often of secondary importance. Consequently, EZ is based on three free parameters, ν , a and T_{er} , to which we add the scaling parameter s fixed to 0.1.

Because of these simplifications, EZ does not require raw data at all, only three descriptive statistics: the mean of all the response times (MRT) in a given condition (which includes successes as well as errors), the variance of those



response times (VRT) and the proportion of correct answers (P_c). This peculiarity makes the EZ model adequate for rough and quick estimations of those three parameters.

Although one of the statistics provided to EZ is commonly called percent correct (P_c),¹ it is important to realize that P_c must truly be the probability of crossing the top boundary; generally, it is assumed that the top boundary is the boundary that must be crossed for a response to be correct. EZ does not know if responses are correct or wrong. It only knows if a boundary was reached. Hence, P_c should more adequately be approached as the "percent top boundary crossing" (P_{top}). Whenever P_{top} is larger than 50%, it implies that the top boundary was reached more often than the lower boundary and consequently, v is a positive value, i.e., it has a slope pointing up. This is a reminder that boundaries must be labelled consistently. In a "Yes"- "No" experiment for example, the top boundary could be the "Yes" boundary. Whenever the correct answer is "Yes", the drift rate should be positive. When "No" is the correct answer, the drift rate should be negative. In this case, the P_{top} should be 1 the observed accuracy so that v points downward to the "No" boundary. For example, if a participant responded "No" correctly in 95% of the trials, then 95% of the "No" responses should be coded as "0" to indicate that the response has accurately crossed the bottom boundary.

The model's weakness –few parameters can be estimated from EZ– is also its strength. Indeed, the fact that fewer parameters are evaluated avoids over-fitting the peculiarities of a given dataset, and consequently, the estimates are more stable across samples. In Ratcliff (2008), simulated responses were generated using the four EZ parameters (see later for the algorithm). Afterwards, either EZ or the full diffusion model were used to estimate the true parameter values. The results showed that the parameter estimates were more precise when EZ was used than when the full diffusion model was used, with estimates having smaller bias and smaller variance. As Wagenmakers et al. (2007) mention, this makes it a more practical and simpler tool for routine analyses of common psychology experiments. According to the authors' tests, the model is quite reliable and gives estimates highly correlated to the diffusion model estimates when tested on a perceptual discrimination task data.

Deriving EZ

EZ, like the diffusion model, is formalized using a differential equation describing the probability of reaching a certain state x at time t . The fundamental equation is (Gar-

diner, 2004)

$$\partial_t p(x, t|z, 0) = v \partial_z p(x, t|z, 0) + \frac{1}{2} s^2 \partial_z^2 p(x, t|z, 0),$$

where $p(x, t|z, 0)$, the probability that the process reaches a count of x at time t given that it started with a count of z at time zero (Wagenmakers, Grasman, & Molenaar, 2005), is a function of v , the mean drift rate (if greater than zero, it is a drift toward the upper boundary) and s^2 , the within-trial variance in drift rate. From this partial differential equation, it is shown that the statistical moments of order n , M_n , of the time t to reach either boundary given a starting point z ($0 < z < a$) follows a recurrent equation given by

$$v \partial_z M_n(z) + \frac{1}{2} s^2 \partial_z^2 M_n(z) = -n M_{n-1}(z),$$

with $M_0(z) = 1$. The boundary conditions, $M_n(0) = M_n(a) = 0$ for any $n \geq 0$, are derived from the fact that if the accumulation starts at zero or starts at a , then it is stopped immediately as one boundary (the top or the bottom respectively) has been reached. As an example, the mean—the first statistical moment—can be expressed from the $n = 1$ moment, yielding the partial differential equation:

$$v \partial_z M_1(z) + \frac{1}{2} s^2 \partial_z^2 M_1(z) = -1. \quad (1)$$

Because this equation is not conditionalized on which boundary was crossed, it implies—as said in the above subsection—that MRT is the mean of all the response times (irrespective of whether they are correct or erroneous). This system, based on Eq. (1) and the boundary conditions, can be solved to find $M_1(z)$ by most symbolic algebra software (e.g., *Mathematica*, see "DerivationOfEZ.nb", a notebook available on the journal's web site, or *Maple*, see Wagenmakers' web site, <http://www.ejwagenmakers.com/papers.html>). Once the mean decision time is found, add T_{er} to get the predicted mean response time (MRT):

$$\text{MRT} = M_1(z) + T_{er}. \quad (2)$$

Likewise, the second statistical moment $M_2(z)$ is found by solving

$$v \partial_z M_2(z) + \frac{1}{2} s^2 \partial_z^2 M_2(z) = -2 M_1(z), \quad (3)$$

from which variance can be found using the well-known relation

$$\text{Var}(z) = M_2(z) - M_1^2(z). \quad (4)$$

In the solutions to Eqs. (1) and (3), by replacing z with $a/2$, we obtain two equations that are based on a , v and s .

¹But note that in the R package EZ2 described later, it is the percent error, P_e , that must be provided



For example, using the shortcut: $y = av/s^2$, and setting z to $a/2$, we find for the first and second moments and the variance

$$M_1(a/2) = \frac{a}{2\nu} \times \frac{e^y - 1}{e^y + 1} \tag{1b}$$

$$M_2(a/2) = \frac{a}{2\nu^3} \times \frac{s^2 (e^y - e^{-y}) + a \left(\left(\frac{\nu}{2} (e^{-y} + e^y) - 3 \right) \right)}{(e^{-y/2} + e^{y/2})^2} \tag{3b}$$

$$\text{Var}(a/2) = \frac{a}{2\nu^3} \times \frac{(s^2 (e^{2y} - 1) - 2av e^y)}{(e^y + 1)^2} \tag{4b}$$

Finally, the boundary position a can be estimated from accuracy with (Ratcliff & Tuerlinckx, 2002; Bussemeyer, 1982)

$$P_T = 1 - \frac{\exp\left(\frac{-2av}{s^2}\right) - \exp\left(\frac{-2zv}{s^2}\right)}{\exp\left(\frac{-2av}{s^2}\right) - 1}, \tag{5a}$$

where P_T is the percent correct, so that in the specific case where $z = a/2$, Eq. 5a simplifies to

$$a = \frac{s^2}{v} \log\left(\frac{P_T}{1 - P_T}\right). \tag{5b}$$

Taken together, Eqs. (2), (4) and (5b) are three equations with four unknowns, s , a , v and T_{er} . It is customary to set s to 0.1 evidence per square root of second. The other three unknowns can be solved, v first (from the response time variance and percent correct, Eqs. 4 and 5a), a in second (from v and percent correct, Eq. 5a), and lastly T_{er} (from mean response time, Eq. 2, and the two other parameters). Thus, we have these three steps, letting $l_T = \log(P_T/(1 - P_T))$ as a shortcut,

Step 1:

$$v = s^2 \sqrt{\frac{l_T(P_T^2 l_T - P_T l_T + P_T - 1/2)}{VRT}}$$

Step 2:

$$a = \frac{s^2}{v} l_T$$

Step 3:

$$T_{er} = MRT - \frac{a}{2v} \frac{1 - \exp(-v a/s^2)}{1 + \exp(-v a/s^2)}$$

and make v negative when P_T is smaller than 0.5.

Note that these equations can be solved for any z , not just when $z = a/2$ (more on this in a later subsection). Also, note the strictly additive role of s^2 in the differential equations. This has implications for simulations as described in a subsequent section on simulating EZ. Finally, note that instead of fixing s to a conventional level, we could derive the third statistical moment which is related to skewness (Rose & Smith, 2001). However, as skewness is difficult to estimate from response time data (and requires larger samples), it is better to consider only the first two statistical moments. As an exercise, we derive here the predicted value of skewness, often noted γ_1 , based on the first three moments, from the relation

$$\gamma_1 = \frac{M_3(a/2) - 3M_1(a/2)\text{Var}(a/2) - M_1^3}{\text{Var}(a/2)^{3/2}}$$

Solving the problem with *Mathematica*, we get

$$\gamma_1 = - \frac{\sqrt{2}a \left(2a^2\nu^2 e^y (e^y - 1) + 6av s^2 e^y (e^y + 1) - 3s^4 (e^y - 1) (e^y + 1)^2 \right)}{\nu^5 \left(-\frac{a(2av e^y - s^2(e^{2y} - 1))}{\nu^3} \right)^{3/2}}$$

As seen in the Application section, skewness will reveal a weakness of EZ.

Units of measurement

In Wagenmakers et al. (2007), the examples are expressed with seconds. However, many prefer to work with milliseconds. In this case, it is necessary that the parameter s be converted to milliseconds; how to do so requires some understanding of the measurement units. The parameter a is a count of evidence and can be expressed by a unit that

we could label e . The drift rate is an average number of evidence per unit of time. Its unit is consequently e / s . The base response time is expressed in seconds. Finally, regarding variability in drift rate s , this variance represents *variance* in e per unit of time, thus e^2/s . Yet, s is generally provided as a standard deviation; its unit is consequently evidence per square root of time or e/\sqrt{s} . The issue of scaling has been previously addressed by Smith, Ratcliff, and McKoon (2014, footnote 5). If it is preferred to express the parameters in milliseconds (ms) instead of seconds (s), and



based on the conversion factor $1\text{ s} = 1000\text{ ms}$, v (in e/s) becomes $v/1000$ (in e/ms), a stays the same, T_{er} (in seconds) becomes $1000 T_{er}$ (in ms) and s ($\text{e}/\sqrt{\text{s}}$) becomes $s/\sqrt{1000}$ (in $\text{e}/\sqrt{\text{ms}}$). We can check this in *R* (see below for explanations on these instructions) with statistics first expressed in seconds:

```
Data2EZ (MRT = .517, VRT = .024, Pc =
.953, s = 0.1)
# $v = 0.2513 $a = 0.1197 $Ter = 0.3012,
```

then in milliseconds:

```
Data2EZ (MRT = 517, VRT = 24000, Pc =
.953, s = 0.1/sqrt(1000) )$; $ )
# $v = 0.002513 $a = 0.1197 $Ter = 301.2.
```

As seen, the estimates are the same.

gEZ: When the starting point is different from $a/2$

As seen in an earlier subsection, the constraint that z is the midpoint between the top and bottom boundary (the latter fixed at zero) is added in the *Mathematical* derivation after the main equations are solved. Hence, nothing prevents varying the starting point parameter. For any z given as a fraction of a , the equations can be solved. Therefore, a generalized version of EZ in which z is fixed *a priori* to a given fraction of a is always possible and results in an analytical solution. For example, letting z be replaced by $a/3$ in Eq. (5a) results in

$$a = \frac{3s^2}{2v} \times \log \left(\frac{P_T - P_T^2 - \sqrt{(4 - 3P_T)P_T^3} + \sqrt{(4 - 3P_T)P_T}}{2(P_T - 1)^2} \right) \tag{5c}$$

The equations tend to be much longer, so it is better to rely on a symbolic manipulation software such as *Mathematica*.

A situation where this would be useful is with regard to a system that needs to scan information in a near exhaustive way. If the amount of information to be processed for a "top" response is doubled whereas the evidence required making a "bottom" response is approximately constant, it makes sense for the top boundary to be located twice as high. In that scenario, z is no longer the midpoint but rather located at one-third the value of a . Likewise, if the information for "top" responses triples, then z is now a quarter of the value of a , etc. We will use this fact to analyze the "Same"- "Different" data in the last section.

Running EZ analyses

EZ can be programmed in many environments. Here we rapidly review two existing ones, EZ2 for *R*, and an *Excel*

spreadsheet. We also describe two new implementations, one for *SPSS* and another one for *Mathematica*. Each environment has its advantages and disadvantages; the choice only depends on the user's preference and familiarity with the selected environment –there exist no inherent disadvantages of using one environment over another.

Using the EZ Diffusion Model in R

There is a package for EZ to be used in *R*, the open source statistical computing software (*R* Core Team, 2016). The package EZ2 (not to be confused with the EZ package for ANOVAs) is maintained by Grasman (2007).

To use the software:

1. You first need to upload the library (this step is to be done only once on a given installation of *R*):

```
install.packages("EZ2", repos = http://
/R-Forge.R-project.org)
```

2. Before using it, load the library EZ2 into memory with:

```
library(EZ2)
```

Note that *R* is case-sensitive; also, line feeds can be inserted in a command between tokens; we use that feature in the subsequent steps to make the command's content easier to read.

3. To get parameters from descriptive statistics (mean RT, noted MRT, variance in RT, noted VRT, and percent top responses, note ambiguously Pc), use:

```
Data2EZ (
Pc = .953,
MRT = .517,
VRT = .024)
```

4. The reverse operation, to get descriptive statistics from parameters, is done with the following:

```
EZ2.mrt (
nu = 0.25133,
a = 0.11974,
Ter = 0.30118,
z = 0.11974 / 2 )
```

where the parameter nu stands for v . In both functions, the value of s can be omitted; its default value is 0.1.

With this command, we choose the statistic to be reported. Here is shown `mrt` (lower case in the EZ2 library), but it could also be `vrt` or `pe` by replacing `mrt` by these words in the above instruction (for these two parameters, do not provide `Ter` as it is unneeded).



Using the EZ Diffusion Model in Excel

The EZ2 package available in R is also available in an Excel spreadsheet. This spreadsheet is available for Excel version 97 and above from <http://raoul.socsci.uva.nl/EZ2/>.

Instructions on how to use the software:

1. Download the Excel spreadsheet EZ2.xls
2. Open the Excel workbook; activate the macros and accept the warning messages. There should be three sheets. The first one is for basic estimations of MRT, VRT, and Pe (truly percent bottom) values from the EZ parameters. Alternatively, you also have access to CVRT and CMRT which are the variance and the mean RT for trials with correct responses only.

Note that this Excel spreadsheet has functions to estimate MRT, VRT, and Pe values from the EZ parameters only. However, it does not accomplish the opposite process of estimating EZ parameters from these data. This can be realized through a parameter search involving the Solver, which must be installed from the Excel installation program. Details on how to do so are provided on the first sheet of the spreadsheet.

Using the EZ Diffusion Model in Mathematica

The EZ diffusion model is available in a package for Mathematica (version 7 or above). This package allows the user to estimate all of the parameters' values from descriptive statistics, or vice versa. It can also be used to simulate trials in full details. It comes in one package, "EZ4Math.m". The file "UsingEZ4Math.nb" which illustrates the use of the package is also attached to this manuscript.

Instructions on how to use the software:

1. Upload the library from this journal's web site and place the file "EZ4Math.m" in any folder.
2. Load the library into memory using:

```
Needs ["EZ4Math`", "location"]
```

where location is the path to the folder where you placed the file "EZ4Math.m" on your computer, between double quote signs. If the backslash character is required on your operating system, double it, e.g. "C:\\user\\Me\\EZ4Math.m". Note that Mathematica, like R, is case sensitive and extra lines can be added anywhere between tokens.

3. To get parameters from descriptive statistics MRT, VRT and Pc, use for example:

```
EZ4Math[ Stats2Param, {MRT, VRT, Pc}, 0.1]
```

in which Pc is truly percent top, and 0.1 is the default value of the scaling factor s. If the default is correct; it can be omitted in all the instructions of this section.

For example:

```
EZ4Math[ Stats2Param, {0.517, 0.024, .953}, 0.1]
```

You can also add units, for example, if you are working in seconds s and counting evidence e:

```
EZ4Math[ Stats2Param, {0.517 s, 0.024 s^2, .953}, 0.1 e/Sqrt[s]]
```

or in milliseconds:

```
EZ4Math[ Stats2Param, {517 ms, 24000 ms^2, .953}, 0.1 e/Sqrt[1000 ms]]
```

In both cases, the estimated threshold is the same 0.119 e, and the parameters Ter and v are scaled by a conversion factor of 1000, as expected.

4. The reverse operation, to get predicted statistics from parameters, is achieved with:

```
EZ4Math[ Param2Stats, {v, a, ter}, s]
```

For example with George's parameters found in Wagenmakers et al. (2007):

```
EZ4Math[ Param2Stats, {0.2513, 0.1197, 0.3012}, 0.1]
```

Using the EZ Diffusion Model in SPSS

To use EZ in SPSS requires one to enter the required statistics (in columns that have to be called MRT, VRT and Pc) for each participant in the data editor, one information per column and one participant per line. Once data is entered, the syntax of Listing 1 at the end of the article can then be executed; the three parameters will appear in 3 additional columns next to each participant' statistics.

Afterwards, the parameters can be plotted or analyzed in the same way as regular descriptive statistics. This method is used in the Application presented in the last section. It can be incorporated to any SPSS syntax (script) after the descriptive statistics are calculated from raw data. Outputting the parameters in different forms such as graphs or tables can then be directly executed in the same SPSS syntax.

In the next section, we simulate an EZ process to obtain simulated responses from a fictitious character created in Wagenmakers et al. (2007) called "George". Listing 2 opens this file containing raw data and aggregate them to get the needed descriptive statistics. You then simply need to run Listing 1.

Simulating an EZ diffusion model

As EZ is fully analytical, there is no reason to simulate the whole process. However, if one wishes to see examples of



the dynamics of evidence accumulation (as illustrated in Figure 1), it is necessary to simulate a diffusion process. Because EZ assumes continuous time steps, very small-time steps must be used in simulations, or equivalently, many time steps per unit of time must be used. Using the second formulation let Δ_t be the number of time steps per unit of time (second is assumed in Wagenmakers et al., 2007). The Algorithm 1 assumes that the parameters ν , a , T_{er} and s are given. As seen from Algorithm 1, the only source of variability in EZ is the moment-by-moment variation in drift.

This algorithm was implemented in the *Mathematica* package “EZ4Math.m” to provide the ability to visualize the evidence count as a function of time for one trial. It is used with:

```
EZ4Math[ Simulation, {v, a, ter}, s,  $\Delta_t$ ,
  nsim]
```

in which Δ_t is the number of steps per unit of time (Δ_t should at least be 1000, the default) and `nsim` is the number of trials to simulate (default to 1000 simulated trials). It returns replacement rules containing “ALLRT”, “ALLSuccess” and “OnePath”. As an example, we used

```
EZ4Math[ Simulation, {0.2513, 0.1197,
  0.3012}, 0.1, 1000, 125]
```

to generate raw data that could have been George’s results. We ran it a few times until the simulated responses’ statistics were very close to the ones given in Wagenmakers et al. (2007). See Table 2.

The replacement rule “OnePath” contains the evidence counts for the last simulated trial only. For example:

```
P = "OnePath" /. EZ4Math [
  Simulation,
  {0.2513, 0.1197, 0.3012},
  0.1, 1000, 1
];
ListPlot [P, Joined -> True ]
```

illustrates one trial’s evidence count.

If one gathers all the states and plot them as a function of step number, one trajectory will emerge. Figure 1 was generated with five calls to the above function. To estimate mean, variance and percent correct, run this simulation many times (at the very least 10,000 times).

As can be seen in Eqs. 1 and 3, the impact of variance can only overestimate the differential equation. Indeed, in the fundamental equation, the term $s^2/2$ is multiplied by a positive infinitesimal. If that quantity is too large, the summation will overshoot, never undershoot. In the present simulation function, when there is not a sufficient number of time steps per unit of time, the mean RT is biased, as it can only be larger than predicted by solving the

equations. Table 3 shows the estimated statistics as a function of Δ_t using George’s parameters from Wagenmakers et al. (2007): drift rate $\nu = 0.2513$ e/s, boundary separation $a = 0.1197$ e, and non-decision time $T_{er} = 0.3012$ s (assuming that $s = 0.1e/\sqrt{s}$). As seen, when very few time steps are simulated per unit of time, the predicted mean is much larger than predicted from the analytical solution. The same is true for percent correct. Finally, variance first overshoots then tends to the correct number. However, with 3,125 time steps simulated per second, the predictions are accurate to two significant digits. Hence, using 1,000 time steps per unit of time can be considered a minimal requirement when simulating EZ or any (continuous) diffusion model.

How good is a fit of EZ to data?

One frequently heard question is how to assess the goodness of fit of EZ. Comparing model fits is a common endeavor and to that end, getting some measure of goodness of fit is a prerequisite.

Technically speaking, the fit of EZ to *descriptive statistics* is always perfect. It takes 3 descriptive statistics and returns three estimates. There is no parameter search and no index of fit at that level. It explains why it is possible to reverse the estimation process retroactively to find the statistics back from a given parameter set.

Although EZ provides a perfect fit at the level of descriptive statistics, there is no guarantee that the fit is perfect at the raw data’s level. When raw data are available (RTs and response selected), then it is possible that an EZ fit may not be perfect or even good. For example, the distribution inferred from the proposed estimates might not match the RT skewness.

One method proposed by Karayanidis et al. (2009) is to qualitatively judge EZ’s fit by comparing it to an Ex-Gaussian fit. This is however only a visual aid and assumes that the Ex-Gaussian is an adequate reference, a fact that is not universally accepted.

Alternatively, when the raw data are available, it is possible to compute an EZ fit. The approach consists in computing the log-likelihood index of fit. This measure can be used to compare nested models (for example, if one parameter is constrained, or if additional parameters are added) or unrelated models. Typical tests of fit based on likelihood includes the likelihood ratio test (an exact test for nested models; Wilks, 1938) or AIC, BIC and the like (for non-nested or unrelated models). See Hélie (2006) and Cousineau and Allen (2015) for more on this topic.

Computing a likelihood is easy when the probability density function (PDF) predicted by a model is available. Fortunately, this function is known for the diffusion model (with no parameter variability as is the case with EZ). In-



Algorithm 1: Simulating a response time and a decision (top or bottom, respectively 1 or 0) assuming a diffusion model with no variability parameters, given the parameters v , a and T_{er} as well as a scaling parameter s and the number of time step per unit of time Δ_t . In the algorithm, italics denotes a comment and the notation $\sim N(\mu, \sigma)$ indicates one realization of a normal distribution with mean parameter μ and scale parameter σ .

```

Before start, no step has been simulated and starting point is midway between 0 and a;
step ← 0;
currentstate ← a/2;
Repeat as long as we are within the boundaries;;
while 0 < currentstate < a do
  step ← step + 1;
  drift ~ N(v/Δt, s/sqrt(Δt));
  currentstate ← currentstate + drift;
end
Determine if the response was actually "top" or "bottom";;
if currentstate > a then
  response ← 1;
else
  response ← 0;
end
RT is the number of steps divided by the number of steps per unit of time, to which Ter is added;
RT ← step/Δt + Ter;
return RT. response

```

deed, the fundamental equation can be solved to find the cumulative distribution function (see Abbasi, 2017; Gardiner, 2004). The formula can also be found in Tuerlinckx

(2004, Eq. 1); an alternative expression is found in Van Zandt, Colonius, and Proctor (2000). Both the PDF and the CDF are given here.

$$f(t | \nu, a, z, ter, s) = \frac{\pi s^2}{a^2} e^{-\frac{\nu z}{s^2}} \sum_{k=1}^{\infty} k \sin\left(\frac{\pi k z}{a}\right) e^{-\frac{1}{2}(t-ter)\left(\frac{\pi^2 k^2 s^2}{a^2} + \frac{\nu^2}{s^2}\right)},$$

$$F(t | \nu, a, z, ter, s) = \frac{e^{-\frac{2a\nu}{s^2}} - e^{-\frac{2\nu z}{s^2}}}{e^{-\frac{2a\nu}{s^2}} - 1} - \frac{\pi s^2}{a^2} e^{-\frac{\nu z}{s^2}} \sum_{k=1}^{\infty} \frac{2k \sin\left(\frac{\pi k z}{a}\right) e^{-\frac{1}{2}(t-ter)\left(\frac{\pi^2 k^2 s^2}{a^2} + \frac{\nu^2}{s^2}\right)}}{\frac{\pi^2 k^2 s^2}{a^2} + \frac{\nu^2}{s^2}}$$

The exact computation of these formulas is not possible as it contains an infinite sum. However, summing a few hundreds of terms is generally enough to get a precise estimate. In breakthrough work, Navarro and Fuss (2009) demonstrated how to determine the number of terms to sum to get a desired precision and which of the two alternative PDF equations (the one above or the one provided in Van Zandt et al. (2000) needs the smallest amount of iteration. They also provided an R function that performs the computations. Thus, it is possible to quantify the PDF and consequently, the likelihood, in a very efficient way (see also Van Zandt, 2000; Cousineau, Brown, & Heathcote,

2004, for more on likelihood functions).

Finally, note that—again—measurement units have an impact on the measure of fit. When RTs are expressed in milliseconds, the time unit is stretched by a factor of 1000. Because the area under the density of the RT distribution must nonetheless be 1, it implies that the height of the PDF is likewise divided by 1000. Thus, the log-likelihood, the sum of the log densities, is reduced by $\log(1000)$ for every datum when these data are expressed in milliseconds relative to when they are expressed in seconds.

We added to the *Mathematica* library an additional function that returns the log likelihood index of fit for a



Table 2 ■ George’s data. These are 125 fictitious data generated so as to match very closely the descriptive statistics of the character George described in (Wagenmakers, Van Der Maas, & Grasman, 2007). The first entry contains a response time (in second), the second entry contains the response made (0 is bottom boundary; 1 is top boundary). The data have been sorted by RT.

Table with 8 columns of numerical data representing response times and responses for 125 entries.

Table 3 ■ Predicted mean and variance in RT as well as proportion correct as a function of the number of simulated time steps per second. In the table, each individual statistic represents the average over 10,000 simulations. As seen, with few time steps per unit of time, the predicted performance is distant from the true performance seen on the last line.

Table with 4 columns: Δt, Mean RT, Variance in RT, Proportion correct. Rows show values for Δt = 1, 5, 25, 125, 625, 3125, ∞.

given set of parameters and a given data set. It is necessary that raw data be available. To test this function, we use again what might have been George’s data, if they had been available, given in Table 2.

Computing the log likelihood index of fit (the sum of the log density for each of the data points; see Van Zandt, 2000; Cousineau et al., 2004) is achieved in the Mathematica “EZ4Math.m” package with:

```
LogLikelihood[
  EZDistribution[{v, A, ter} ],
  Data
]
```

where v, A, ter are the parameters estimated and data is a list of pairs RT, response. If the scaling parameter is different from the default 0.1, add its value past the curly braces with

```
LogLikelihood[
  EZDistribution[{v, A, ter}, s ],
  Data
]
```

Here is an example with George’s data evaluated at the position of the estimated George’s parameter

```
LogLikelihood[
  EZDistribution[{ 0.2513, 0.1197,
  0.3012} ],
  GeorgeData
]
```

It computes the fit of George’s parameters to George’ data generated earlier (and provided in the Mathematica package under the variable "GeorgeData" for easy access).

When performed, the fit of George’s data is 64.3461



(generally, log likelihood fits are negative numbers, but it does not have to be the case when PDF are narrow, extending above 1). This is an *Excellent* fit, but not a surprising fit as the data were generated using Algorithm 1. To see the match between predicted and observed distributions, Figure 2 shows the results for both the cumulative distribution function (CDF, left panels) and the probability density function (PDF, right panels) for the top boundary (top panels) and the bottom boundary (bottom panels) responses; the data are shown using histograms and the model using lines. Both superimpose almost perfectly. Note that if the data and parameters are expressed in ms, the fit is -799.1239; we can verify that $64.34614 - 125 \times \log(1000)$ is indeed -799.1239.

Such an index of fit can be used to compare models. For example, the full diffusion model can be compared to EZ. When the additional parameters η , s_t and s_z (which are zero in EZ) are given a very small value (e.g., 0.00001), the fit of the full diffusion model is 64.3927, a very small improvement in fit (of 0.0466, not significant, $p > 0.999$ according to a likelihood ratio test, $\chi^2(3) = 0.0932$). However, if the parameter s_z is set beyond zero to 0.06 or larger, then the fit of the full diffusion model decreases to 62.3341 or below, a significant decrement ($\chi^2(3) = 4.0244$, $p = 0.045$).

If we run a full diffusion model parameter estimate analysis (minimizing the log-likelihood index of fit), we get those seven parameter estimates: $\nu = 0.2669$, $\eta = 0.0037$, $z = 0.0550$, $s_z = 0.00001$, $T_{er} = 0.2900$, $s_t = 0.0022$ and $a = 0.1215$. The index of fit is 65.2768, a fit slightly better than the fit assessed by EZ. However, when penalized for model complexity using AIC, the fit is 2.1159 points in favor of the simpler EZ model. As seen, the common parameters ν , a and T_{er} are quite comparable between the EZ fit and the DM fit. The starting point z is also found to be roughly halfway through 0 and a .

An application with "Same"- "Different" Data

To illustrate the use of EZ, we chose a dataset from our laboratory. It bears on the "Same"- "Different" task in which participants must indicate as fast as possible and with as few errors as possible whether two stimuli presented in close succession are fully identical or not. The stimuli were strings of consonants of various lengths; 50% of the trials show pairs of identical strings. These strings can differ on as few as a single letter or as many as be composed of entirely different letters. The experiment is similar to Bamber's seminal research (1969); however, we added case manipulations: On half of the trials, the letter-case of the first and second stimuli could be mismatching; the participant had to make their decision solely on the identity of the letter, irrespective of the stimuli's case.

The data features a *fast*- "Same" effect whereby "Same" responses tend to be much faster than "Different" responses for a given string length. This effect is still not entirely explained (see Farell, 1985; Krueger & Shapiro, 1981; Proctor, 1981; Sternberg, 1998). There has been a previous attempt to fit the data with a diffusion model (Ratcliff, 1985) which was not successful and criticized (Proctor, 1986). This section describes the experiment and the results of the fit. Not surprisingly, because EZ is a nested model within the diffusion model, it will not fit better the results, but its failure will shed some light on future modeling directions.

Method

Participants

Twenty individuals participated in the experiment. They were Canadian residents, all students at the University of Ottawa. They were aged 18 or over; no additional demographic information was collected. Their vision was normal or corrected to normal. They spoke French or English.

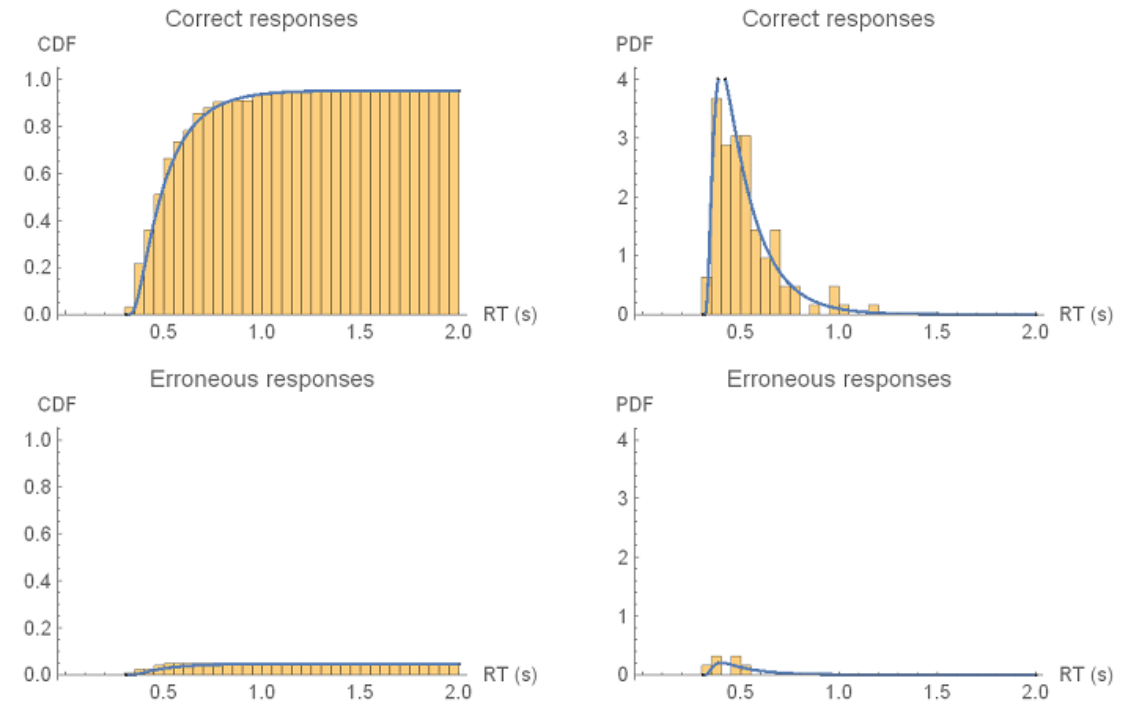
Procedure

The participant's role was to determine whether a first stimulus (noted S_1) is "Same" or "Different" from a second stimulus (S_2). The stimuli are strings of letters containing from one to four letters. Conditions are balanced so that there are as many "Same" trials as "Different" trials and that each length of strings are presented equally (S_1 and S_2 are of the same length). For each trial in the experiment, there is a blank screen for 500 ms, a fixation point for 500 ms, then S_1 is presented for 400 ms, a second blank of 400 ms, followed by S_2 , which stays until the participant answers or for 5000 ms. The letters used in the strings were the same as the ones used by Bamber (1969): B, C, D, F, J, K, L, N, S, T, V, or Z; they never appeared more than once in a stimulus.

The complexity (the total number of letters) of the stimuli varied from 1 to 4 letters through the experiment, and when the strings were different, the number of differences varied from a single letter different to all letters different. The new manipulation of this experiment is a change in casing: half the strings were entirely lowercase and the other half entirely uppercase so that on half of the trials, S_1 and S_2 are presented with a different case. Hereafter, we call "mismatching trials" those where the case was different in S_2 relative to S_1 . The goal of the experiment was to determine the impact of the letter's casing on the participant's performance in a "Same"- "Different" task. The participants completed 768 trials in a 55-minutes session; all the conditions were presented in a random order.



Figure 2 ■ Fit of the EZ parameters to George’s simulated data of Table 2. The left panels show the density of the data (bars) and the predicted distribution (line) for the correct (top) and error (bottom) responses. The distributions are defective ones, so that they do not totalize 1 but to the proportion of responses of a given type (95.3% correct on top, 4.7% of error on bottom). The fit found by EZ is very close to perfect for these data that were generated using EZ.



Results

We collected 15360 total measures. We checked the data for outliers. Analyses of the full diffusion model require that extreme data be removed because their associated probability are very low and results in underflow in many programming environments. We consequently used the same stringent approach here. We have found 46 RTs above 3000 ms, but 40 of them are from a single participant (participant 14). Regarding low outliers, we found 58 RTs below 220 ms, but 46 are from a single participant (participant 12). We chose to exclude RTs below 220 ms and above 3000 ms and also removed the participants 12 and 14. We are left with 13,806 data, or 99.6% of the data from the 18 remaining participants.

In order to simplify the presentation of the results, we examined whether there was any differences in mean RTs between the order of case (was S_1 in lower case and S_2 in upper case different from S_1 in upper case and S_2 in lower case?) and found no significant differences ($F(1, 502) = 0.243, p = 0.622$). We also examined whether there was a difference between cases in which the stimuli were both

uppercase or both lowercase and found no significant differences either ($F(1, 502) = 0.107, p = 0.743$). Further, as was observed in Harding (2018), for the “Different” responses, there was no difference between matching cases and mismatching cases ($F(1, 358) = 2.332, p = 0.128$). For the “Same” trials however, there was an important effect of matching vs. mismatching cases ($F(1, 142) = 7.508, p = 0.007$). These results were entirely predictable from the facilitation framework of Proctor (1981) and replicated many times with various manipulations (Harding, 2018). Hereafter, to facilitate presentation of the results, we collapsed the “Different” trials from matching and mismatching case conditions but kept “Same” trials separated.

The results of the experiment are shown in Figure 3 where we see mean RTs. We also show the standard deviations of RTs, averaged across participants, as the variance of RTs is used by EZ. The “Same”-“Different” task typically returns rather constant standard deviation across experimental conditions (Harding, 2018). In the third panel, we show percent correct averaged across participants for each condition. Finally, we show the mean skewness of RTs. The results show that the “Same” trials were faster on average



in the matching case condition than in the mismatching case condition (the blue lines in each panel). Proportion of correct answers gets lower for trials with higher complexity and lower number of differences (and RTs get slower, so this is not a speed-accuracy trade-off). The results concur with results from previous experiments (e. g., Bamber, 1969).

We estimated the EZ parameters using the SPSS syntax (see Listing 1). We assumed that the "Same" response was emitted through the top boundary and the "Different" responses with the bottom boundary. The parameters were then averaged across participants; see Figure 4 for the results.

As we can see, the drift rate ν seems fairly constant when the two stimuli are "Same" despite the fact that the length of strings is increasing. This would imply that additional matching letters do not increase processing noise and consequently, do not decrease processing rate. On the other hand, the drift rate becomes slower (closer to zero) when at least one letter of the stimuli is different, especially when there is only one difference. This result reflects the fact that RTs are longer in those trials and suggests that these trials have less salient information or more noise hindering processing rate, potentially affected by the number of letters, as more complexity induces more noise in the decision-making process.

Mean boundary separation seems to be decreasing, although this trend is fairly weak given the wide variations in the estimates (as evidenced by the wide error bars). This result for "Same" trials goes against expectations: The "Same" trials should involve some form of exhaustiveness (all letters must be checked for matches before a "Same" response can be made); higher level of exhaustiveness can only be achieved by increasing the distance between starting point and boundary position. Thus, the "Same" response line (blue) should behave quite differently from the "Different" response lines, something completely missing from the present results which may be attributable to the fact that the starting position in EZ is fixed to the mid-point between 0 and a in every condition.

T_{er} estimates account for the majority of the variance of RT results (correlation is 0.845), which is counterintuitive: If most of the changes in response times are not caused by decisional processes, then it is difficult to explain why participants are slower for more complex stimuli. As seen in Figure 4, there are significant effects between conditions for T_{er} : The one-difference trials take longer base times compared with trials having more differences, as is the case of mean RTs. This is problematic because it implies that most of the response time is spent on encoding and motor response times. The magnitude of T_{er} is often larger than many of the fast response times, something that

should be impossible from the diffusion model. On average 1.2% of the trials produced responses that are faster than the estimated T_{er} parameter. In the condition with 4 letters, only 1 of which is mismatching, this proportion rises to 5.5%. Thus, EZ seems to wrongfully estimate the processes surrounding the accumulation of evidence in this specific experiment.

To determine if it is because EZ has too many constraints that it cannot capture the results in a meaningful way, we ran two exploratory analyses using variations of EZ. First, we note that in EZ, internal noise is fixed. Yet, with additional letters, it might be possible that within-trial randomness is increased, a possibility that Ratcliff and Hacker (1981) alluded to in their Experiment 1. To accommodate internal noise, we chose to set the within-trial variability s to 0.1 times the total number of letters. The rationale for doing this is that with more letters, there should be a larger amount of noise in the decision process. This is done in EZ by increasing s . This variation is easily implemented in the SPSS syntax by changing the line

```
Compute s = 0.1.
```

to

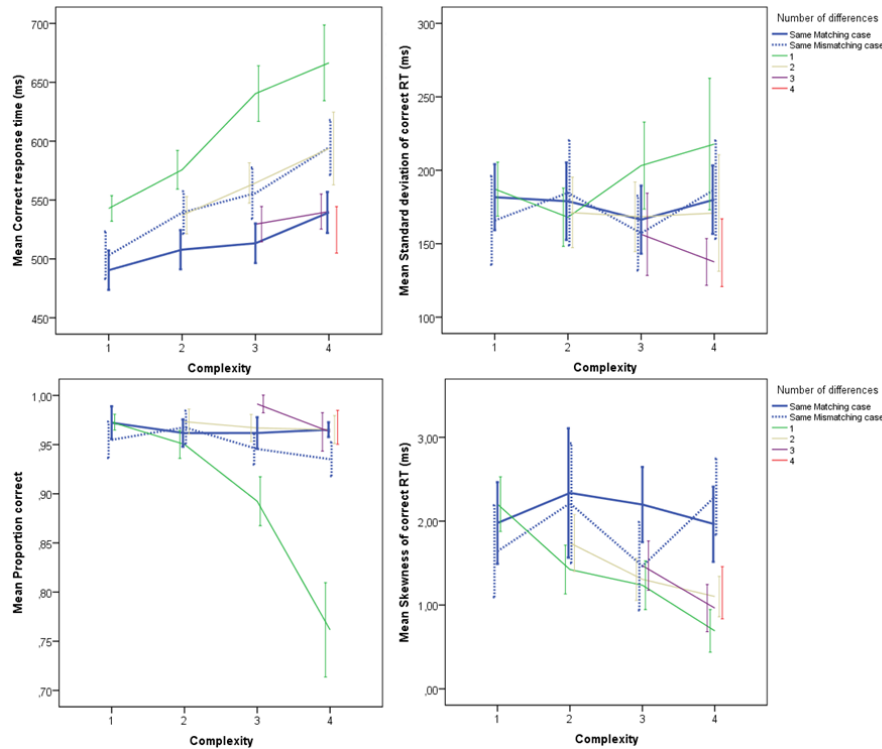
```
Compute s = 0.1 * ntotal.
```

where $ntotal$ is the SPSS variable containing the number of letters in the string. The results were however not convincing for three reasons. First, EZ suggests that the mean separation between boundaries is increasing for both "Same" and "Different" responses. Yet, for "Different" trials, a single difference is enough to trigger a response and consequently, "Different" threshold should not be increasing. Second, drift rates sped up with an increasing number of letters, a result quite contrary to expectation. Finally, we again have the problematic result that most of RT variations are attributable to base response times T_{er} . About 1.1% of the RTs are below the lower bound predicted by T_{er} . This percentage rises to 4.6% when the complexity is 4 and the number of differences is 1. Because non-decisional time effects bear no explanation regarding the changes in RTs in the "Same"-"Different" task, we are still left in an inconclusive state.

In a second variation, we decided to change the starting point of the model (the parameter z) so that it is located at $a/(n_{total} + 1)$. The rationale for doing this is that "Same" responses (the top boundary) should be exhaustive to detect all-identical letters, so that the boundary must move further away from the starting point when the number of letters increases. The "Different" responses, on the other hand, are self-terminating: any single mismatch is enough, irrespective of the total number of letters in the string. Hence, the distance between the lower boundary



Figure 3 ■ Mean response times, mean standard deviations, proportion correct, and mean skewness as a function of the number of differences (ndiff) and the total number of letters (ntotal) (for the "Same" responses, estimated separately for the matching case and mismatching case conditions) from the data of the "Same" - "Different" experiment. Error bars shown are correlation-adjusted 95% confidence intervals of the mean.



and the starting point should be a constant whereas the distance between the starting point and the upper boundary must increase as the total number of letters increases. This variation can only be performed in *Mathematica* by solving Eqs. 1b, 3b and 5b with $a/2$ replaced by $a/(n_{total} + 1)$. Note that with EZ, we cannot force the z_s to have the same value across conditions, only that its separation to the top boundary is proportional to the number of letters. The results of this new set of analyses exacerbated the problem regarding T_{er} : According to this version of EZ, about 400 ms is dedicated to base response time in the 1-difference "Different" trials. Yet, more than 20% of the trials are completed prior to 400 ms, this figure is just not reasonable, and we have to conclude that this version of EZ is providing an inadequate fit to the data.

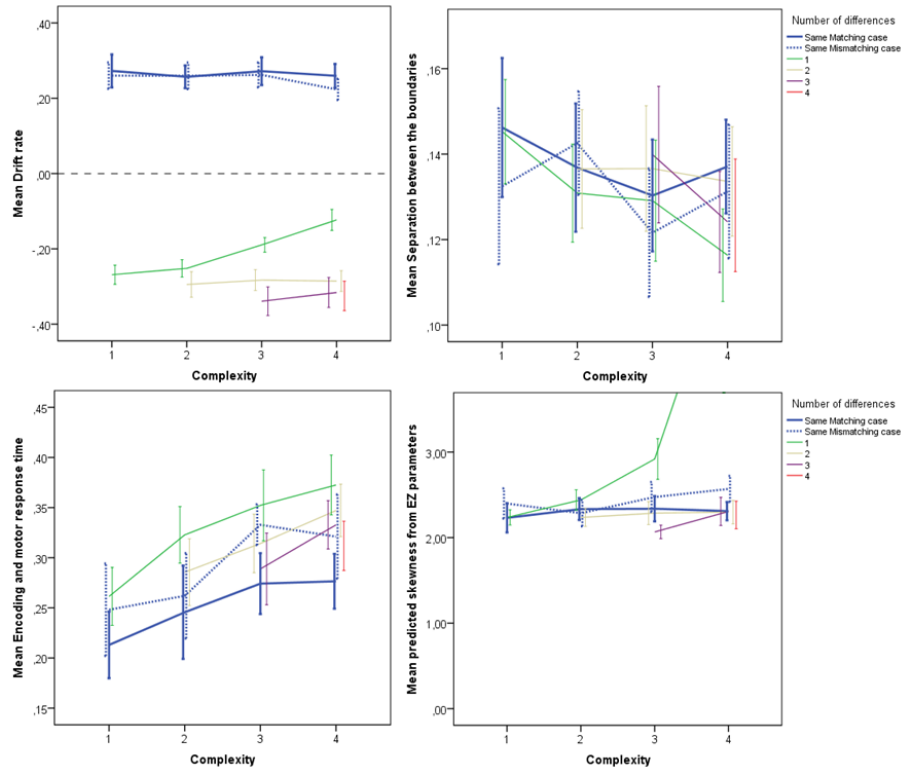
Discussion

To understand the inability of EZ to fit the data, and following Wagenmakers et al. (2008), we ran three checks. First, we checked the shape of the RT distribution. For a data to be a good fit, the RTs must be skewed, which is the case

here. Also, the spread should become more pronounced as task difficulty increased. However, this is not what is observed in the current experiment. If we examine Figure 3, we see mean RTs increasing with the number of letters, but this increase is not mirrored by an increase in standard deviations. A possibility is that the participants want to reduce response times in the slower "Different" condition by truncating the slow responses. That predicts more errors in the "Different" condition, as is observed in the results. Second, we checked the relative speed of error responses. For a good fit with EZ, correct and error response RTs should have the same distribution because both responses have the same (null) variability parameters s_t and s_z . This was not found in the results: error response times seem to have the same means irrespective of the response whereas the correct RTs are faster for the "Same" than for the "Different" responses (the fast-same effect). There are very few errors in most conditions, so it is difficult to ascertain the replicability of this result (see e.g., Krueger & Shapiro, 1981). Third, we checked for a bias in the starting point by plotting the mean RTs to visually judge whether



Figure 4 ■ Mean drift rate (v), mean separation between boundaries (a), mean encoding and motor response time (T_{er}), and predicted skewness from these parameters as a function of ndiff, and ntotal (for the "Same" responses, estimated separately for the matching case and mismatching case conditions) with parameter s set to $0.1e/\sqrt{s}$. Error bars show 95% confidence intervals of the mean.



the means cross over in such a way that errors are fast for one response and slow for another response. This was not the case: as seen above, the errors were flat whereas there was a “fast-same” effect for the correct responses. Yet, Ratcliff and Hacker (1981) argued strongly for a decision bias in the “Same”-“Different” task.

In our opinion, the most notable discrepancy between the model and the data is to be found in the skewness of the RTs. In the empirical data, as seen in Figure 3, skewness is below 2; there is even a trend for skewness to be decreasing for more difficult “Different” conditions. On the other hand, when we look at the predicted skewness from EZ, they are either stable just above 2 or increasing in the condition with the slowest responses and more errors. Looking back at Figure 1, we see that if the drift rate is smaller, the arrow pointing towards the upper boundary is less steep. Consequently, the RTs are spread out towards very long response times, increasing the spread and/or the skewness. Yet, none of these trends are seen in the data. Quite the contrary: Conditions with longer response times

are not more spread out and are not more skewed. This may explain why T_{er} is overly large: this parameter compensates for the inability of the model to fit data that are slower but not more skewed and more spread out.

General discussion

EZ diffusion model is an attempt to create an easier and more accessible version of the diffusion model. The result is a simpler, computationally cheaper and easily manipulated estimation procedure. However, by doing so, it loses the full generalizability of the Ratcliff’s diffusion model, which contains more parameters but can fit the entire RT distribution, not just three descriptive statistics summarizing the raw data. EZ can be used with many software (with two new additions described herein) including *R*, *Excel*, *SPSS* and *Mathematica*.

One limitation of EZ is that all the conditions are fitted separately. For example, consider a “Yes”-“No” experiment in which the upper boundary triggers a “Yes” response and the lower boundary, a “No” response. The same a param-



eter should be used in both conditions. However, it is not possible with EZ to constraint parameters to be equal between conditions. This is the role of a design matrix as the one implemented in DMAT (Vandekerckhove & Tuerlinckx, 2008).

In the same vein, the core assumption behind EZ, the nullification of across trial variability, has been criticized by Ratcliff and Smith (2015): “[...] it is unreasonable to assume that subjects can set their processing components to identical values on every equivalent trial of an experiment [...]” (p. 7).

We tested EZ from a dataset obtained using a “Same”-“Different” task. Its utility in this task seems however questionable, as the results suggested that EZ attributed most of the response time to non-decisional processes (as estimated by T_{er}). This result is counterintuitive as the motor responses are identical between conditions and the perception processes should be fairly equivalent in seeing “Same” strings and seeing “Different” strings. Also, on many trials, the observed RTs were smaller than the estimated parameter T_{er} which is a lower bound in the diffusion model. Thus, the estimates proposed by EZ and the variations we tested are just not realistic. We were able to locate the inability of EZ to fit the present results to a problem of skewness. EZ has difficulties predicting more symmetrical distributions when difficulty increases. Yet, this is what was observed in the data. Wagenmakers et al. (2007) noted that this is something that needs to be checked in the data before proceeding to an EZ analysis. Nonetheless, the present results do not endanger the model; rather, it shows that with regards to the “Same”-“Different” task, an adequate model is still to be found.

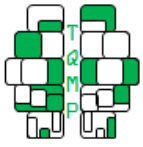
One core assumption underlying diffusion models is the single evidence count. Alternative models such as accumulator models (LBA in particular; Brown & Heathcote, 2008) are similar but rejected this central assumption by postulating that on every trial, multiple counts are active in parallel. Each of these counts can trigger a different response. How well it would fit the “Same”-“Different” data remains to be seen.

References

- Abbasi, N. (2017). Analytical solution to diffusion-convection pde in 1d. Retrieved from http://12000.org/my_notes/diffusion_convection_PDE/diffusion_convection_PDE.htm
- Bamber, D. (1969). Reaction times and error rates for “same” - “different” judgments of multidimensional stimuli. *Perception & Psychophysics*, 6, 169–174. doi:10.3758/BF03210087
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178. doi:10.1016/j.cogpsych.2007.12.002
- Bussemeyer, J. R. (1982). Choice behavior in a sequential decision-making task. *Organizational Behavior and Human Performance*, 29, 175–207. doi:10.1016/0030-5073(82)90255-0
- Cousineau, D. (2009). Fitting the three-parameter weibull distribution: Review and evaluation of existing and new methods. *IEEE Transactions on Dielectrics and Electrical Insulation*, 16, 281–288. doi:10.1109/TDEI.2009.4784578
- Cousineau, D., & Allen, T. A. (2015). Likelihood and its use in parameter estimation and model comparison. *Mesure et Évaluation en Éducation*, 37, 63–98.
- Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments, & Computers*, 36, 742–756. doi:10.3758/BF03206555
- Cousineau, D., Harding, B., Thivierge, J. P., & Lacouture, Y. (2015). Constructing a group distribution from individual distributions using geometric quantile averaging. *Canadian Journal of Experimental Psychology*, 70, 253–277. doi:10.1037/cep0000069
- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin and Review*, 18(1), 61–69. doi:10.3758/s13423-010-0022-4
- Farrell, B. (1985). Same. -“Different” judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, 98, 419–456. doi:10.1037/0033-2909.98.3.419
- Gardiner, C. W. (2004). *Handbook of stochastic methods (3rd ed.)* Berlin: Springer.
- Grasman, R. (2007). Ez2: A package for the ez diffusion model (Version 1.2.0).
- Harding, B. (2018). *A single process of the same-different task* (Doctoral dissertation, Unpublished).
- Hélie, S. (2006). An introduction to model selections: Tools and algorithms. *Tutorials in Quantitative Methods for Psychology*, 2, 1–10. doi:10.20982/tqmp.02.1.p001
- Karayanidis, F., Mansfield, E. L., Galloway, K. L., Smith, J. L., Provost, A., & Heathcote, A. (2009). Anticipatory reconfiguration elicited by fully and partially informative cues that validly predict a switch in task. *Cognitive, Affective, & Behavioral Neuroscience*, 9(2), 202–215. doi:10.3758/CABN.9.2.202
- Krueger, L. E., & Shapiro, R. G. (1981). A reformulation of proctor’s unified theory for matching-task phenomena. *Psychological Review*, 88, 573–581. doi:10.1037/0033-295X.88.6.573



- Lerche, V., & Voss, A. (2018). Speed-accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates? *Behavior Research Methods, Online*, 50(6), 2568–2585. doi:[10.3758/s13428-018-1034-7](https://doi.org/10.3758/s13428-018-1034-7)
- Luce, R. D. (1986). *Response times, their role in inferring elementary mental organization*. Oxford: Oxford University Press.
- Navarro, D. J., & Fuss, I. J. (2009). Fast and accurate calculations for first-passage times in wiener diffusion models. *Journal of Mathematical Psychology*, 53(4), 222–230. doi:[10.1016/j.jmp.2009.02.003](https://doi.org/10.1016/j.jmp.2009.02.003)
- Proctor, R. W. (1981). A unified theory for matching-task phenomena. *Psychological Review*, 88, 291–326. doi:[10.1037/0033-295X.88.4.291](https://doi.org/10.1037/0033-295X.88.4.291)
- Proctor, R. W. (1986). Response bias, criteria settings, and the fast-same phenomenon: A reply to ratcliff. *Psychological Review*, 93, 473–477. doi:[10.1037/0033-295X.93.4.473](https://doi.org/10.1037/0033-295X.93.4.473)
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from. Retrieved from <http://www.R-project.org/>
- Ratcliff, R. (1978). A theory of memory retrieval. the oxford handbook of computational and mathematical psychology. (edited by busemeyer, j. R. et al.). 85(2), 59–108. doi:[10.1037/0033-295X.85.2.59](https://doi.org/10.1037/0033-295X.85.2.59)
- Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, 92, 212–225. doi:[10.3758 / BF03204200](https://doi.org/10.3758/BF03204200)
- Ratcliff, R. (2008). The ez diffusion method: Too ez? *Psychonomic Bulletin & Review*, 15, 1218–1228. doi:[10.3758/PBR.15.6.1218](https://doi.org/10.3758/PBR.15.6.1218)
- Ratcliff, R., & Hacker, M. J. (1981). Speed and accuracy of same and different responses in perceptual matching. *Perception and Psychophysics*, 30, 303–307. doi:[10.3758/BF03214286](https://doi.org/10.3758/BF03214286)
- Ratcliff, R., & Smith, P. (2015). Modelling simple decisions and applications using a diffusion model. In J. R. e. a. Busemeyer (Ed.), *The oxford handbook of computational and mathematical psychology* (pp. 1–51). doi:[10.1093/oxfordhb/9780199957996.013.3](https://doi.org/10.1093/oxfordhb/9780199957996.013.3)
- 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(3), 438–481. doi:[10.3758/BF03196302](https://doi.org/10.3758/BF03196302)
- Röhner, J., & Thoss, P. (2018). Ez: An easy way to conduct a more fine-grained analysis of faked and non-faked iat data. *The Quantitative Methods for Psychology*, 14(1), 17–37. doi:[10.20982/tqmp.14.1.p017](https://doi.org/10.20982/tqmp.14.1.p017)
- Rose, C., & Smith, M. D. (2001). *Mathematical statistics with mathematica*. New York: Springer-Verlag.
- Smith, P. L., Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on jones and dzhafarov (2014). *Psychological Review*, 121(4), 679–688. doi:[10.1037/a0037667](https://doi.org/10.1037/a0037667)
- Smith, P., & Ratcliff, R. (2015). Diffusion and random walk processes. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (pp. 395–401). doi:[10.1016/B978-0-08-097086-8.43037-0](https://doi.org/10.1016/B978-0-08-097086-8.43037-0)
- Sternberg, S. (1998). Inferring mental operations from reaction-time data: How we compare objects? In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Methods, models, and conceptual issues* (pp. 365–445). Cambridge, Mass: Bradford Book.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, England: 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. doi:[10.3758/BF03206552](https://doi.org/10.3758/BF03206552)
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7, 424–465. doi:[10.3758/BF03214357](https://doi.org/10.3758/BF03214357)
- 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. doi:[10.3758/BF03212980](https://doi.org/10.3758/BF03212980)
- van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: Ez, fast-dm, and dmat. *Journal of Mathematical Psychology*, 53(6), 463–473. doi:[10.1016/j.jmp.2009.09.004](https://doi.org/10.1016/j.jmp.2009.09.004)
- Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with matlab: A dmat primer. *Behavioral Research Methods*, 40(1), 61–72. doi:[10.3758 / BRM.40.1.61](https://doi.org/10.3758/BRM.40.1.61)
- Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, 16, 44–62. doi:[10.1037/a0021765](https://doi.org/10.1037/a0021765)
- Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, 39, 767–775. doi:[10.3758/BF03192967](https://doi.org/10.3758/BF03192967)
- Wagenmakers, E.-J., Grasman, R. P. P. P., & Molenaar, 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. doi:[10.1016/j.jmp.2005.02.003](https://doi.org/10.1016/j.jmp.2005.02.003)



- Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, 14(1), 3–22. doi:[10.3758/BF03194023](https://doi.org/10.3758/BF03194023)
- Wagenmakers, E.-J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. (2008). Ez does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review*, 15, 1229–1235. doi:[10.3758/PBR.15.6.1229](https://doi.org/10.3758/PBR.15.6.1229)
- Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). Hddm: Hierarchical Bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, 7, 1–20. doi:[10.3389/fninf.2013.00014](https://doi.org/10.3389/fninf.2013.00014)
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9, 60–62. doi:[10.1214/aoms/1177732360](https://doi.org/10.1214/aoms/1177732360)

Open practices

📄 The *Open Material* badge was earned because supplementary material(s) are available on [the journal's web site](#).

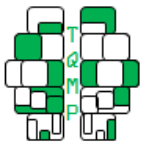
Citation

T. Groulx, J., Harding, B., & Cousineau, D. (2020). The EZ diffusion model: An overview with derivation, software, and an application to the Same-Different task. *The Quantitative Methods for Psychology*, 16(2), 154–174. doi:[10.20982/tqmp.16.2.p154](https://doi.org/10.20982/tqmp.16.2.p154)

Copyright © 2020, T. Groulx, Harding, and Cousineau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 30/07/2018 ~ Accepted: 20/08/2019

Listings 1 and 2 follow.



Listing 1 ■ SPSS syntax to get v , a and T_{er} from the columns MRT, VRT and Pc assuming that $s = 0.1$. Lines beginning with asterisks are comments. This can be copy-pasted in a syntax file.

```
* s is the scaling factor parameter, fixed.
COMPUTE s = 0.1.
* if you are using ms, divide 0.1 by sqrt(1000):.
* COMPUTE s = 0.1 / sqrt(1000).

* temporary variable.
COMPUTE logitpc = ln(Pc/(1-Pc)).

* the rate parameter.
DO IF Pc < 1/2.
COMPUTE v= -s*sqrt(sqrt((logitpc*(pc**2 * logitpc - Pc*logitpc + Pc - 1/2)) /VRT)).
ELSE.
COMPUTE v= +s*sqrt(sqrt((logitpc*(pc**2 * logitpc - Pc*logitpc + Pc - 1/2)) /VRT)).
END IF.

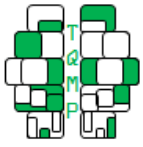
* the boundary separation.
COMPUTE a = s**2 * logitpc / v.

* a shortcut variable.
COMPUTE y = - v * a / s**2.

* the estimated mean decision time.
COMPUTE mul = (a / (2 * v )) * ((1-Exp(y)) / (1 + Exp(y) )).

* the encoding + motor response time.
COMPUTE Ter = MRT - mul.

* some cleaning.
VARIABLE LABELS v "drift rate".
VARIABLE LABELS a "Separation between the boundaries".
VARIABLE LABELS Ter "encoding + motor response time".
FORMAT v (F8.3) a (F8.3) Ter (F8.3).
DELETE VARIABLES y logitpc mul.
EXECUTE.
```



Listing 2 ■ SPSS commands to open a raw data file, and generate the needed descriptive statistics so that Listing 1 is ready to be run.

```
SET DECIMAL DOT.  
  
* set the working directory and open the file.  
CD "C:\\Users\\Me\\Documents".  
  
DATA LIST list file = "GeorgeData.dat"  
  /RT rep.  
  
* reduce data to mean, standard deviation and percent correct.  
AGGREGATE outfile = "GeorgeStatistics.SAV"  
  /MRT = Mean (RT)  
  /SDRT = SD (RT)  
  /Pc= Mean (rep).  
  
* get the obtained aggregates and transform sd in variance.  
GET file = "GeorgeStatistics.SAV".  
COMPUTE VRT = SDRT**2.  
EXECUTE.
```