Psychological Review
2007, Vol. 114, No. 3, 830–841


0033-295X/07/$12.00 DOI: 10.1037/0033-295X.114.3.830

# THEORETICAL NOTE

# On the Linear Relation Between the Mean and the Standard Deviation of a Response Time Distribution

Eric-Jan Wagenmakers
University of Amsterdam

Scott Brown
University of Newcastle

Although it is generally accepted that the spread of a response time (RT) distribution increases with the mean, the precise nature of this relation remains relatively unexplored. The authors show that in several descriptive RT distributions, the standard deviation increases linearly with the mean. Results from a wide range of tasks from different experimental paradigms support a linear relation between RT mean and RT standard deviation. Both R. Ratcliff's (1978) diffusion model and G. D. Logan's (1988) instance theory of automatization provide explanations for this linear relation. The authors identify and discuss 3 specific boundary conditions for the linear law to hold. The law constrains RT models and supports the use of the coefficient of variation to (a) compare variability while controlling for differences in baseline speed of processing and (b) assess whether changes in performance with practice are due to quantitative speedup or qualitative reorganization.

*Keywords:* response times, diffusion model, variability, law

In many psychological experiments, the efficiency of mental processing is quantified by response time (RT; Laming, 1968; Link, 1992; Luce, 1986; Townsend & Ashby, 1983). For instance, in the lexical decision task, high-frequency words (e.g., *SMOKE*) are classified faster than are low-frequency words (e.g., *FUME*), supporting the claim that in lexical decision, high-frequency words are processed more efficiently than are low-frequency words. Because RT is both informative and easy to measure, it has become one of the most important dependent variables in psychological research. Attention has traditionally centered on changes in mean RT across experimental conditions. In separate literatures on aging and practice effects, the focus has widened to include RT variability as a useful indicator of cognitive performance (for aging, see Hultsch, MacDonald, & Dixon, 2002; Li, 2002; MacDonald, Hultsch, & Dixon, 2003; Shammi, Bosman, & Stuss, 1998; for practice, see Heathcote, Brown, & Mewhort, 2000; Logan, 1988, 1992). In mathematical psychology, several models have been extended to make detailed predictions regarding the shape of the entire RT distribution (e.g., Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Brown & Heathcote, 2005; Logan, 1988, 1992; Ratcliff & Smith, 2004; Usher & McClelland, 2001).

Empirical evidence and theoretical considerations point to several general characteristics of RT distributions in psychological tasks (cf. Ratcliff, 2002). First, RT distributions are decidedly nonnormal—they are almost always skewed to the right. Second, this skew increases with task difficulty. Third, the spread of the distribution increases with the mean. Although the above characteristics are so general as to invite the term "law," surprisingly little work has been done to quantify the regularities. An exception is Luce's (1986, p. 64) analysis of Chocholle's (1940) data, in which he noted that the standard deviation of RT was remarkably linear in mean RT for a signal-detection experiment. In this article, we attempt to sharpen the third law of RT distributions in psychology; that is, we aim to quantify the precise empirical relation between RT mean and RT variability.

This research was inspired by a recent analysis of the diffusion model, a successful model for how people make speeded decisions (e.g., Ratcliff, 1978, 2002). The diffusion model predicts that when task difficulty increases, RT mean and RT standard deviation increase at the same rate (Wagenmakers, Grasman, & Molenaar, 2005). That is, the diffusion model predicts that the relation between RT mean and RT standard deviation is linear. The present work began as a test of this prediction.

The outline of this article is as follows. The first section shows that four out of five commonly used descriptive RT distributions naturally accommodate a linear relationship between RT mean and RT standard deviation. The second section shows that data from a wide range of experimental paradigms support the assertion that the relation between RT mean and RT standard deviation is linear. The third section provides an explanation of the linear relationship in terms of the diffusion model (cf. Wagenmakers et al., 2005), and the fourth section provides an explanation of the linear relationship

Eric-Jan Wagenmakers, Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands; Scott Brown, School of Psychology, University of Newcastle, Newcastle, New South Wales, Australia.

Preparation of this article was supported by the Australian Research Council Linkage International Project LX0348125 and a Veni grant from the Netherlands Organisation for Scientific Research. Both authors believe the other did most of the work. We thank Tim Rickard and Gordon Logan for helpful comments on an earlier version of the article.

Correspondence concerning this article should be addressed to Eric-Jan Wagenmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. E-mail: EJ.Wagenmakers@gmail.com

in terms of the instance theory of automatization (e.g., Logan, 1995). The fifth section discusses three situations for which the linear law has been shown to fail—these are all situations in which the diffusion model is known to be inappropriate. The sixth section discusses the theoretical and practical implications of the strong linear relation between RT mean and RT standard deviation, and the seventh section concludes the article.

## Descriptive RT Distributions

A descriptive RT distribution allows researchers to succinctly describe the observed data. The estimated parameters of the descriptive distribution may be used to quantify and test the effects of experimental manipulations (cf. Hervey et al., 2006; Rouder, Lu, Speckman, Sun, & Jiang, 2005). Several descriptive distributions are in use, and all of them accommodate the universal right skew in RT data. Here, we consider the ex-Gaussian, the shifted lognormal, the shifted Wald, the shifted Weibull, and the Gumbel distribution (see, e.g., Heathcote, Brown, & Cousineau, 2004).

Interest centers on the nature of the relation between mean and standard deviation as a function of task difficulty. Table 1 summarizes the results of our analyses (see Appendix A for details). The ex-Gaussian, shifted lognormal, shifted Weibull, and Gumbel distributions all provide natural accounts of a linear relationship between standard deviation and mean RT, provided certain parameters are used to index task difficulty. For each distribution, the parameter that ensures a linear relationship if it indexes task difficulty is $\tau$ for the ex-Gaussian, $\mu$ for the shifted lognormal, $\tau$ for the shifted Weibull distribution, and $\sigma$ for the Gumbel distribution. The shifted Wald distribution is the only descriptive RT distribution that does not accommodate a linear relationship between mean and standard deviation for any simple parameter changes.

A corollary of our result is the inverse interpretation of the prior paragraph: When considering parameter estimates from the four descriptive RT distributions that generate a linear relation between RT mean and RT standard deviation, the parameters mentioned may be interpreted as estimates of task difficulty. Another corollary of our result is that when a researcher reports certain parameters as estimates of task difficulty, he or she implicitly acknowledges that the relation between RT mean and RT standard deviation is linear.

## Empirical Evidence

### *Analysis 1: A Single Experiment*

We reanalyzed data from a study by Brown, Marley, Donkin, and Heathcote (2006). Their experiment was well suited to the current task: There were 28 separate within-subject conditions, and sufficient data were collected from each participant to avoid averaging data across participants. The experimental task was absolute identification. Participants were shown eight lines of different lengths—the shortest line was called Line 1, the longest line was called Line 8. On each trial, participants were shown one line and asked to decide which of the eight lines it was. Responses were timed accurately using a microphone and voice key. Nine participants each contributed 5,600 responses, divided into 70 blocks of 80 trials each; blocks were alternately assigned to speed or accuracy emphasis conditions. During speed emphasis blocks, participants were instructed to respond as quickly as they could, and during accuracy emphasis blocks, participants were instructed to take their time to ensure a high accuracy rate. Within each of the 35 speed and accuracy emphasis blocks, 20 blocks used all eight lines as stimuli, 10 blocks used only the central four lines (Lines 3, 4, 5, and 6), and 5 blocks used only two lines (Lines 4 and 5). For full details of the experimental design, see Brown et al.'s article.

Brown et al.'s (2006) design resulted in 200 observations from each of 28 within-subject conditions for 9 participants. Below, we analyze those 28 conditions separately, after removing RTs associated with incorrect responses. Calculations of variance are sensitive to outliers, so we censored RTs greater than 2.5 s, eliminating 2.9% of the overall data. The exclusion of these outlier RTs did not influence the qualitative pattern of results.

We calculated 28 means and standard deviations for each of the within-subject conditions for each participant. These are shown in Figure 1, with nine panels for the 9 participants (in no particular order). A linear relationship between mean and standard deviation is evident for all participants. Also shown on Figure 1 are lines of best fit, and the $r$ values in the lower right corners show the strength of the linear relationships. The same axis values are used for all plots to aid comparisons.

The $r$ values ranged from .86 to .96, with a mean of .92, indicating a very strong linear relationship between mean and standard deviation (when RTs longer than 2.5 s were included, the mean $r$ was .90). These kinds of high correlations are not common for psychological phenomena; in RT research, they are usually only observed for other laws of RT, such as Fitts's law[1] (Fitts, 1992) or the Hick–Hyman law (Hick, 1952; Hyman, 1953; McMillen & Holmes, 2006). The high correlations are particularly impressive because both correlated variables (mean and standard deviation) were estimated with noise. Other laws, such as the Hick–Hyman law, Fitts's law, or the law of practice, have only one random variable, with the other fixed by design. We would like to point out that the high correlations are not due to the fact that for some participants the data appear to form two clusters (i.e., one for speed emphasis and one for accuracy emphasis): Correlations calculated just within the accuracy emphasis cluster are about as high as those for the total data set.

To give an idea of how strong the linear relationship is, the standard error of prediction, averaged across participants, was just 49 ms for estimating standard deviation from an observed mean RT and just 81 ms for estimating mean RT from standard deviation. This result is quite promising, but of course it is dangerous to base our conclusions on data from just one experiment.

### *Analysis 2: Nine More Experiments*

To explore the generality our findings, we reanalyzed data from a survey of RT experiments used by Heathcote et al. (2000) to investigate changes in RT with practice. The survey included 17 data sets, of which we were able to use only nine (the other data sets had fewer than eight within-subject conditions, making them inappropriate for the analyses below). These nine data sets covered a range of paradigms, from visual search tasks to memory experiments and problem-solving tasks, and included data from 127

[1] For an overview, see http://en.wikipedia.org/wiki/Fitts'_law.

Table 1
*Relation Between the Mean and the Standard Deviation for the Most Popular Descriptive Response Time Distributions*

| Distribution | Relation |
|---|---|
| Ex-Gaussian | |
| Pdf | $f(x) = \frac{1}{\tau\sqrt{2\pi}}\exp\left(\frac{\sigma^2}{2\tau^2} - \frac{x-\mu}{\tau}\right)\int_{-\infty}^{[(x-\mu)/\sigma]-(\sigma/\tau)} \exp\left(-\frac{y^2}{2}\right)dy$ |
| *M* | $\mu + \tau$ |
| *SD* | $\sqrt{\sigma^2 + \tau^2}$ |
| Relation | Linear as a function of $\tau$ if $\tau \gg \sigma$. |
| Shifted lognormal | |
| Pdf | $f(x) = \frac{1}{(x-\theta)\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln(x-\theta)-\mu}{\sigma}\right)^2\right]$ |
| *M* | $\theta + \exp\left(\mu + \frac{1}{2}\sigma^2\right)$ |
| *SD* | $\sqrt{\exp[2\mu + \sigma^2](\exp(\sigma^2) - 1)}$ |
| Relation | Linear as a function of $\mu$; linear as a function of $\sigma$ if $\sigma > 2$. |
| Shifted Wald | |
| Pdf | $f(x) = \frac{a}{\sqrt{2\pi(x-\theta)^3}}\exp\left\{-\frac{[a-\mu(x-\theta)]^2}{2(x-\theta)}\right\}$ |
| *M* | $\theta + a/\mu$ |
| *SD* | $\sqrt{a/\mu^3}$ |
| Relation | Nonlinear. |
| Shifted Weibull | |
| Pdf | $f(x) = c\tau^{-c}(x-\theta)^{c-1}\exp\{-[(x-\theta)/\tau]^c\}$ |
| *M* | $\theta + \tau\Gamma(c^{-1}+1)$ |
| *SD* | $\tau[\Gamma(2c^{-1}+1) - \Gamma^2(c^{-1}+1)]^{1/2}$ |
| Relation | Linear as a function of $\tau$. |
| Gumbel | |
| Pdf | $f(x) = \frac{1}{\sigma}\exp\left\{-(x-\mu)/\sigma - \exp\left[\frac{-(x-\mu)}{\sigma}\right]\right\}$ |
| *M* | $\mu + 0.578\sigma$ |
| *SD* | $\sigma\pi/\sqrt{6}$ |
| Relation | Linear as a function of $\sigma$. |

*Note.* See Appendix A for more details regarding the different distributions. Relation refers to the relation between mean and standard deviation. Pdf = probability density function.

participants. We refer to the nine experiments with the alphanumeric codes shown in Table 2. A description of the data sets and references to the original publications can be found in Appendix B (see Heathcote et al., 2000, for more details).

For each of these nine experiments, we split the data according to the within-subject conditions described in Appendix B. Each within-subject condition also included a practice effect condition, which we (arbitrarily) split into four levels. We then proceeded as for the absolute identification experiment above: We calculated the mean and standard deviation for each condition and then calculated the *r* value representing the strength of linear association between mean and standard deviation for each participant. Figure 2 shows box plots of the *r* values obtained for the nine experiments (right-hand side), as well as a box plot of the *r* values taken from Figure 1 (left-hand side) for reference.

Each of the nine experiments presented essentially the same results as our analyses of the absolute identification experiment. The correlations between mean RT and standard deviation of RT

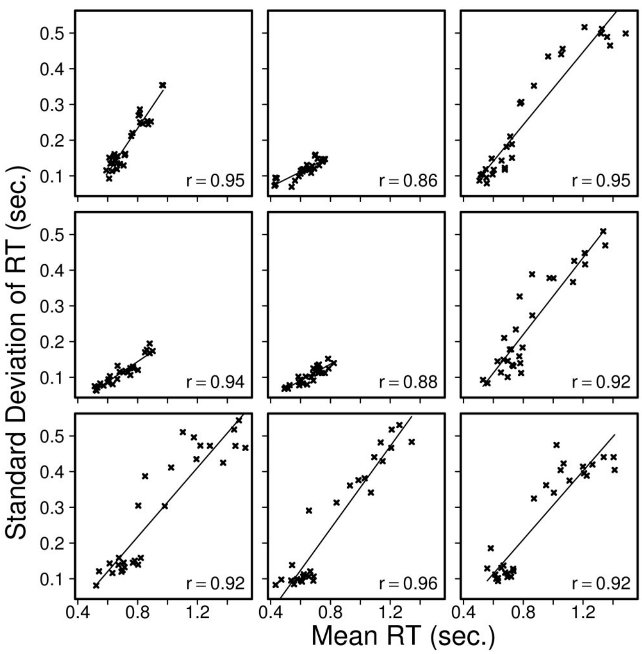


*Figure 1.* Standard deviation of response time (RT) versus mean RT (both in seconds). Each plot represents one participant. There were 28 within-subject conditions, shown by crosses. Also shown are lines of best fit and *r* values for each participant.

were strong for almost all participants. The mean *r* across all participants in all experiments was .87: Only 11 out of the 136 participants showed an $r < .75$, and about two thirds of the participants (99 out of 136) yielded an $r > .85$. The analysis of data that include practice level as a factor is complicated by changes in mean RT within conditions. For example, in the initial stages of practice, RT can change dramatically because of fast improvement. This change can artificially inflate variability estimates, confounding true variability in RT with changes in mean RT. We did not control for this confound in the analyses above. However, we recalculated all analyses using detrended mean RT

Table 2
*Summary of Data Set*

| Source | Data set name | Censor |
|---|---|---|
| Brown et al. (2006) | AbsID | RT < 2.5 |
| Palmeri (1997) | C1, C2, C3 | None |
| Rickard & Bourne (1996) | M1 | 0.2 < RT < 5.0 |
| Rickard (1997) | M2 | None |
| | A1 | 0.2 < RT < 10.0 |
| Heathcote & Mewhort (1993) | V1, V3 | None |
| Carrasco et al. (1998) | V2 | None |

*Note.* See Appendix B for descriptions of the experiments. The censor column presents the criteria used, if any, to censor outliers from the data set. We used censoring specified by the original authors of each data set. Letter and number combinations for data set names refer to type of task and data set number. AbsID = absolute identification; RT = response time; C = counting; M = mental arithmetic; A = alphabetic arithmetic; V = visual search.

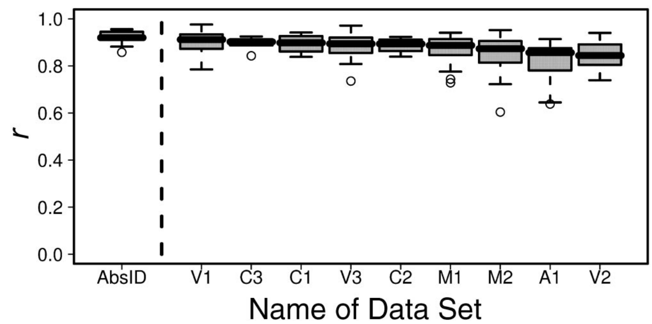


*Figure 2.* Box plots of *r* values for the participants in each of 10 experiments. The left-most box plot represents the nine *r* values from Figure 1 (the absolute identification experiment [AbsID]). The other nine box plots represent similar analyses for nine experiments from other laboratories and other researchers (see Table 2 and Appendix B for further details). Letter and number combinations for data sets refer to type of task and data set number. V = visual search; C = counting; M = mental arithmetic; A = alphabetic arithmetic.

and standard deviation estimates. Detrended estimates were obtained by fitting exponential curves to each practice series (see Heathcote et al., 2000, for methodological details). Then, for each within-subject condition, the ordinary standard deviation used above was replaced by the standard deviation of residuals from the exponential function estimate. This detrending operation produced results only a little weaker than those shown in Figure 2: The mean *r* was .83; only 32 out of the 136 participants yielded an $r < .75$, and 61 out of 136 participants yielded an $r > .85$.

### *Statistical Tests*

To check that the relationship between the mean and standard deviation truly was linear or close to linear, we first applied the Wald–Wolfowitz runs test. This algorithm tests the null hypothesis that the residuals from the linear fit are random by assessing the number of times their signs (positive or negative) alternate. We took the residuals from the linear regression of standard deviation on mean RT for each individual participant in each of the 10 data sets described above. We controlled familywise error by applying Bonferroni's correction within each data set. That is, to keep the chance of making a Type I error at about 5% for each of the 10 data sets, we used a significance level of .05 divided by the number of participants within that data set. The results supported our conclusion that the relationship between mean and standard deviation is linear: Of the 10 data sets, 6 of them never demonstrated a violation of the null hypothesis. Three other data sets (namely, C3, C1, and M1) each yielded just one violation of the null hypothesis (out of 5, 4, and 24 tests, respectively). The remaining data set (V3) showed just two significant tests (out of 16).

As a second statistical check, we tested the null hypothesis of a perfectly linear relation between mean and standard deviation (i.e., $r = 1$). If this null hypothesis is exactly true, one would expect 7 out of the 136 data sets to be rejected at the .05 level. In reality, the null hypothesis was rejected in 15 out of 136 data sets—impressive, given the stringent nature of the test. Out of the 15 data sets for which the perfectly linear relation was rejected, 12 came from data set V3. This data set has a large number of conditions, and this leads to the rejection of the null hypothesis whenever the correlation falls below .92.

### *Comparison Against Alternative Measures of Dispersion*

We further tested the linearity of the relationship between mean and standard deviation of RT by using two other measures of dispersion: variance and interquartile range. We used these two measures to establish whether the linear relationship between mean and standard deviation was somehow privileged or was generally true for any measure of dispersion. For each of the two dispersion measures, we repeated the analyses presented above, arriving at a measure of linearity for each individual participant's data (the correlation, *r*, between the dispersion measure and the mean RT). We compared these correlations with those obtained using the standard deviation, as shown in Table 3. The relationship between the mean and standard deviation is more linear than either of the relationships between the mean and the variance or the mean and interquartile range. This is true in aggregate over the nine experiments and also within each and every individual data set. Further, the mean correlation values are always greater for the relationship between mean and standard deviation than for either of the other two relationships.

In a final analysis, we set out to compare the model that assumes a linear relation between mean and standard deviation against a more general model that can also account for nonlinear relations. In particular, we compared a linear model (L) that describes the data as $SD = a + b \times M$ against an alternative model (N) that adds a parameter $c$ to capture nonlinearities: $SD = a + b \times M^c$. Because the models are nested, the more complex nonlinear Model N will always fit the data better than the linear Model L. The empirical issue is whether the increase in goodness of fit obtained by adding the extra parameter $c$ is large enough to warrant the additional complexity of Model N over Model L (cf. Wagenmakers & Waldorp, 2006). Using classical nested-model tests on the difference in $R^2$, we found that the simpler Model L was preferred over Model N for 133 out of 136 participants. A Bayesian analysis with uninformative priors would produce results that favor Model L still more (see Wagenmakers & Grünwald, 2006, for an illustration). Even though this result speaks against the use of the power function as a relationship between mean and standard deviation, we briefly analyzed the parameter estimates from those functions. Recall that a parameter of $c = 1$ indicates a linear mean–standard deviation relationship and a parameter of $c = 0.5$ indicates an (approximately) linear mean–variance relationship. We observed parameter estimates of $c < 1$ more often than we observed $c > 1$ (92 vs. 44 participants, respectively), but the results were not overwhelming. Overall, these curve-fitting results indicate that Model L is to be preferred over the more complex Model N.

## A Diffusion Model Account of the Linear Relation Between Mean and Standard Deviation

From a theoretical perspective, the relation between RT mean and RT spread has recently been explored using Ratcliff's Wiener diffusion model (e.g., Ratcliff, 1978, 2002). The diffusion model is a sequential sampling model in which noisy evidence is accumulated over time until a prespecified evidence threshold is reached. Figure 3 illustrates the model as applied to the lexical decision task (e.g., Ratcliff, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, & McKoon, in press). The model assumes that the total RT for a particular trial is the sum of a nondecision component of processing (e.g., stimulus encoding and response execution) and a decision component of processing: RT = Nondecision component + decision component.

For our purposes, the nondecision component of processing is assumed to be fixed at some value, $T_{er}$. The decision component of processing is modeled as a continuous random walk (i.e., diffusion) process. The decision process begins at a starting point $z$ in between the response threshold for the "word" and the "nonword" response. A noisy information accumulation process then drives the decision process until it reaches one of the two response boundaries, after which the corresponding response is initiated. Figure 3 shows two example processes that both reach the "word" threshold. The distance between the response boundaries $a$ is a measure for response caution: When the boundaries are set close together, RT will be fast, but, because of the impact of chance fluctuations in the decision process, response accuracy will be low.

Table 3
*Comparison of Linearity of the Relationship Between Mean and Standard Deviation With Mean Versus Variance and Mean Versus Interquartile Range (IQR)*

| | Standard deviation is more linear | | Mean correlation | | |
|---|---|---|---|---|---|
| Data set | Variance | IQR | *SD* | Variance | IQR |
| AbsID | 5/9 | 7/9 | .920 | .914 | .904 |
| C1 | 4/4 | 3/4 | .894 | .821 | .873 |
| C2 | 4/4 | 4/4 | .887 | .834 | .862 |
| C3 | 5/5 | 5/5 | .894 | .817 | .862 |
| M1 | 24/24 | 19/24 | .870 | .799 | .835 |
| M2 | 17/19 | 14/19 | .852 | .814 | .816 |
| A1 | 17/21 | 17/21 | .818 | .793 | .778 |
| V1 | 20/24 | 19/24 | .899 | .885 | .875 |
| V2 | 8/10 | 10/10 | .846 | .828 | .785 |
| V3 | 11/16 | 15/16 | .882 | .862 | .836 |
| Total | 115/136 | 113/136 | .870 | .835 | .836 |

*Note.* Letter and number combinations for data set names refer to type of task and data set number. AbsID = absolute identification; C = counting; M = mental arithmetic; A = alphabetic arithmetic; V = visual search.

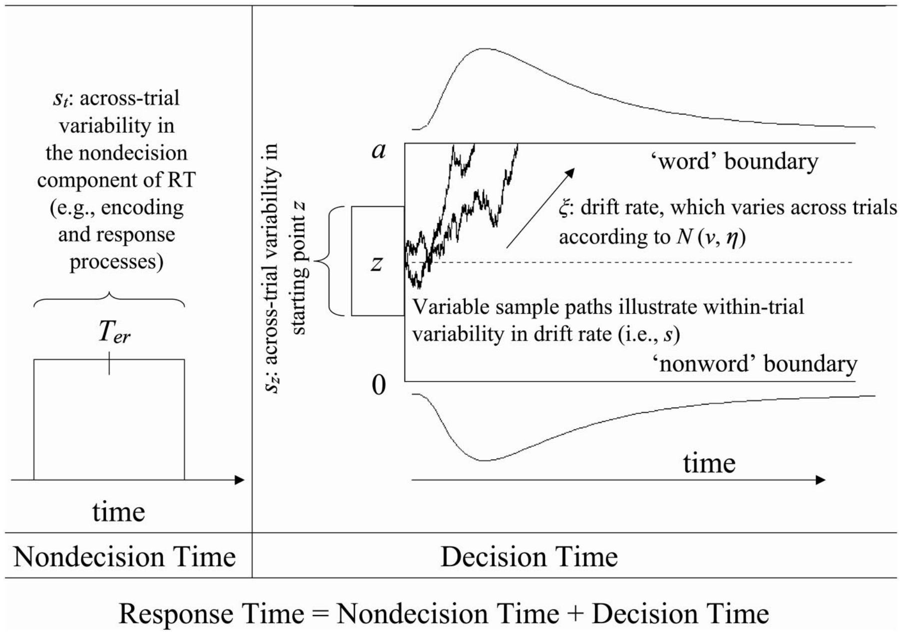


*Figure 3.* Ratcliff's (1978) diffusion model for the lexical decision task. See text for details. RT = response time.

In Ratcliff's diffusion model, task difficulty is quantified by drift rate $v$, which is defined on the real line; $v > 0$ and $v < 0$ lead to evidence accumulation consistent with the "word" and "nonword" response, respectively. When the absolute value of $v$ is high, the impact of chance fluctuations is low, resulting in fast and accurate decisions. Parameter $s$ is a scaling parameter that quantifies the stochastic, nonsystematic component of the information-accumulation process on each trial.

Wagenmakers et al. (2005) studied the predictions of the Ratcliff diffusion model with respect to the relation between RT mean and RT variance. The usual Ratcliff diffusion model includes some extra complications (between-trial variability in three parameters). However, to facilitate the mathematical derivation, a simplified ("EZ") version of the diffusion model was used that corresponds exactly to the above description, with the starting point always in the middle of the response boundaries (i.e., $z = a / 2$; cf. Wagenmakers, van der Maas, & Grasman, in press). In the simplified diffusion model, the mean decision time (MDT) is given by

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

where $y = -va / s^2$. The variance of the decision time (VDT) is given by

$$\text{VDT} = \begin{cases} \left[\frac{a}{2v}\right]\left[\frac{s^2}{v^2}\right]\frac{2y\exp(y)-\exp(2y)+1}{(\exp(y)+1)^2} & \text{if } v \neq 0 \\ \frac{a^4}{24s^4} & \text{if } v = 0 \end{cases}, \tag{2}$$

from which it follows that the relation between MDT and VDT is given by

$$\text{VDT} = \begin{cases} \text{MDT} \times \left[\frac{s^2}{v^2}\right]\frac{\exp(2y)-2y\exp(y)-1}{\exp(2y)-1} & \text{if } v \neq 0 \\ \text{MDT} \times \frac{a^2}{6s^2} & \text{if } v = 0 \end{cases}. \tag{3}$$

Figure 4 plots the standard deviation of decision time (i.e., $\sqrt{\text{VDT}}$) against MDT. Each of the six lines corresponds to a different level of boundary separation $a$; each separate line was constructed by varying drift rate $v$. The values for $a$ (i.e., $a \in [0.07, 0.17]$) and $v$ (i.e., $v \in [0.1, 0.5]$) were chosen to be representative ofvalues encountered in past research with the model (cf. Wagenmakers van der Maas, & Grasman, in press, Figure 2).

As is evident from Equation 3 and Figure 4, the diffusion model predicts that a manipulation of task difficulty (i.e., drift rate) should result in an approximately linear relation between RT mean and RT standard deviation (Wagenmakers et al., 2005).[2] This prediction from the diffusion model has consequences for a large number of models and tasks. For instance, Logan (2002) proposed the instance theory of attention and memory (ITAM), which integrates formal theories of attention, memory, and categorization. In ITAM, the preferred decision rule is instantiated by the random walk model (Logan, 2002, p. 393). As the random walk model is the discrete version of the diffusion model, the result of the above

[2] It is possible for a decrease in boundary separation to compensate for an increase in drift rate in such a way that response accuracy remains constant. This change in boundary separation leaves the approximately linear relation between RT mean and RT standard deviation intact (analyses not shown here for reasons of brevity).

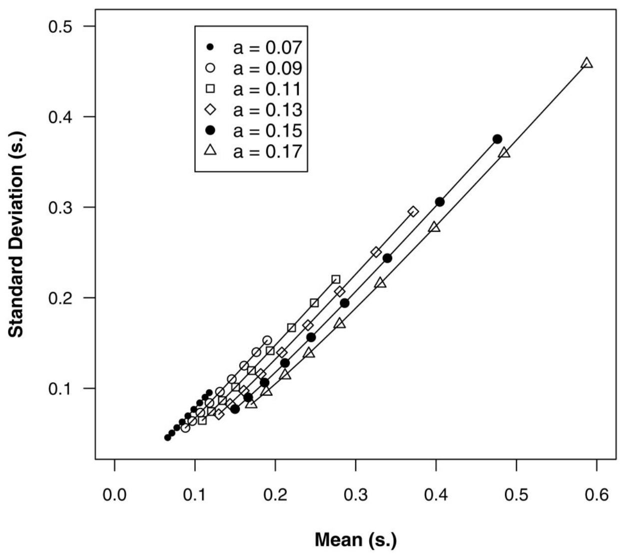


*Figure 4.* Ratcliff's (1978) diffusion model predicts an approximately linear relation between the mean and standard deviation (both in seconds) of decision time when task difficulty (i.e., drift rate) is varied from 0.10 to 0.5 in steps of 0.05 (adapted from Wagenmakers et al., 2005, Figure 3).

analysis has the potential to carry over to ITAM and the tasks and models with which ITAM is associated.

## An Instance Theory Account of the Linear Relation Between Mean and Standard Deviation

Another theoretical perspective on the linear relation between RT mean and RT standard deviation is provided by the instance theory of automatization (ITA; Compton & Logan, 1991; Lassaline & Logan, 1993; Logan, 1988, 1990, 1992, 1998; Logan & Klapp, 1991). The ITA accounts for the effects of practice in terms of memory storage and retrieval. The ITA assumes that, at first, performance on a given task is based exclusively on the speed with which an algorithm is carried out (e.g., solving 3 + 5 by counting on one's fingers). With practice, performance becomes more based on retrieval of the correct answer (i.e., cuing memory with "3 + 5" and retrieving "8"). Performance is said to be automatic when it is based on memory retrieval instead of the execution of an algorithm.

According to the ITA, each encounter with a stimulus results in the storage of a separate memory trace (i.e., an instance). When the same stimulus is presented again in the same task context, all corresponding memory traces race to be retrieved. Each memory trace races independently of the other traces. A response is executed when the first memory trace is retrieved, provided that such retrieval occurs before completion of the algorithm. The more memory traces participate in the race for retrieval, the sooner the first memory trace tends to be retrieved and so the faster the response tends to be. This is the mechanism by which the ITA accounts for the speedup of RT with practice.

According to the ITA, it is not just mean RT that speeds up with practice but it is the entire RT distribution that speeds up. In the model, the distribution of RTs is the distribution of the minimum of *n* memory traces racing for retrieval. It can be shown that under fairly general conditions, the distribution of the minimum tends toward a Weibull distribution as *n* increases (cf. Cousineau, Goodman, & Shiffrin, 2002; for a discussion, see Colonius, 1995; Logan, 1995). With practice (i.e., as *n* increases), the mean and standard deviation of this distribution decrease at the same rate (Logan, 1988). This means that the ITA predicts a linear relation between RT mean and the RT standard deviation as a function of practice. Elements of this prediction were extensively tested and validated by Logan (1992).

## Span of the Linear Law

The linear relationship between RT mean and RT standard deviation does not hold under all circumstances; just as other laws of RT (e.g., Fitts's law and the Hick–Hyman law), it is subject to boundary conditions. As it happens, these boundary conditions are in perfect agreement with Ratcliff's (1978) diffusion model.

### *Boundary Condition 1: Manipulations of Nondecision Time*

In the Ratcliff diffusion model, the linear relation between RT mean and RT standard deviation comes about through a systematic variation of task difficulty (i.e., drift rate). When parameters other than drift rate vary across conditions or participants, the diffusion model does not necessarily predict a linear relation between RT mean and RT standard deviation. For instance, variation in response conservativeness (i.e., boundary separation) does not lead to the linear law (Wagenmakers et al., 2005).

In the context of a model for the psychological refractory period, Sigman and Dehaene (2005, 2006) recently showed

> that the perceptual transformation of sensory information . . . can be carried out in parallel with another task and is a low-variability process (whose variability does not increase with the mean); that the accumulation of evidence establishes a bottleneck and is an intrinsically variable process; and that the execution of the response constitutes yet another parallel, low-variability process. (Sigman & Dehaene, 2005, p. 344)

In terms of the diffusion model, the parallel processes associated with perception and response execution together constitute the nondecision component, whereas the serial bottleneck process associated with the accumulation of evidence constitutes the decision component of processing. With respect to the nondecision time component in the diffusion model, one needs to distinguish between within-trial and across-trial variability. The diffusion model assumes that the within-trial variability of nondecision time is zero; over the course of a single trial, the nondecision component is not associated with a stochastic process. The nondecision component does vary across trials, however. This variability, $s_t$, does not depend on the mean, $T_{er}$, so that experimental manipulations can affect $T_{er}$ without affecting $s_t$. The results from Sigman and Dehaene's (2005) work strongly suggest that this is indeed the case. It should further be noted that the across-trial variability in the nondecision component of processing is a relatively recent addition to the model; in many applications of the model to data, the impact of $s_t$ is not pronounced (cf. Ratcliff & Tuerlinckx, 2002).

### *Boundary Condition 2: Mixtures*

The Ratcliff diffusion model assumes the presence of a single, one-shot underlying cognitive decision process; the model does not predict that the linear law holds when task performance is a mixture of two or more different processes and the mixture proportion changes over time. The model does predict that the linear law holds within each separate mixture component. To illustrate how mixtures can lead to a nonlinear relation between mean and standard deviation, it is useful to consider an example involving the component power law model (CMPL; Rickard, 1997, 1999, 2004; see also Delaney, Reder, Staszewski, & Ritter, 1998).

The CMPL theory proposes that in learning a new task, performance is first based on a relatively slow algorithm. Later in practice, performance is based on a fast retrieval strategy. The execution of both the algorithm and the retrieval strategy speeds up according to a power function. The CMPL theory predicts that as participants gradually transition from the slow algorithm strategy to the fast retrieval strategy, the RT mean decreases but the RT standard deviation first increases and then decreases—this is due to the fact that RTs are a mixture of two processes with different speeds, and the variance of a mixture includes variability from each component in addition to variability due to the difference between the processes' means. Under such a scenario, the relation between RT mean and RT standard deviation need not be linear; Figure 5 shows data reported in Rickard (2004) that clearly violate linearity.[3] Of course, our analyses suggest that when data are separated by strategy (thus removing the mixture property), the relationship between mean RT and standard deviation will, once again, be linear.

### *Boundary Condition 3: Serial Processing*

The third boundary condition on the linear law concerns mental architectures in which processing is serial and exhaustive (for details, see Townsend & Ashby, 1983, pp. 192–201). For example, Sternberg (1966) used an item-recognition task in which participants were presented with short lists of items. After each list, participants were given a probe item to which an "old" or a "new" response was required, depending on whether the probe item was a part of the previous list. Sternberg showed that the mean RT increased linearly with the length of the to-be-remembered list, suggesting that participants compare the list items with the probe item one by one and that each comparison takes a fixed amount of time. Moreover, the increase in mean RT with list length was the same for lists that required an "old" response or a "new" response. These findings suggested that processing in this memory-scanning paradigm is serial and exhaustive (i.e., every item of the list is compared with the probe item) rather than serial and self-terminating (i.e., the comparison process stops after a match is detected between the probe item and a list item).

When processing is serial and exhaustive, it can be shown that under certain assumptions (e.g., independence of completion times for the probe-to-item comparisons) the RT mean will increase linearly with RT variance instead of with RT standard deviation (Townsend & Ashby, 1983). This boundary condition is again consistent with the Ratcliff diffusion model—recall that the diffusion model is appropriate only in the presence of a single, one-shot underlying cognitive decision process; the diffusion model is inappropriate in the presence of a process that is serial and exhaustive.

The different predictions of the two models suggest that a test to adjudicate between a serial exhaustive processing account and a diffusion model account is to see whether RT mean is linear with RT variance or with RT standard deviation. In the data sets under consideration, serial exhaustive processing is a plausible explanation of performance in the alphabet arithmetic task (A1) and in the visual search tasks (V1, V2, and V3). In every one of these data sets, well over half of the participants had more linear relationships between RT mean and RT standard deviation than between RT mean and RT variance. Across all four tasks, 56 out of 71 participants showed a more linear relationship between RT mean and RT standard deviation than between RT mean and RT variance. For the data under consideration here, a diffusion account appears to be more plausible than a serial exhaustive processing account.

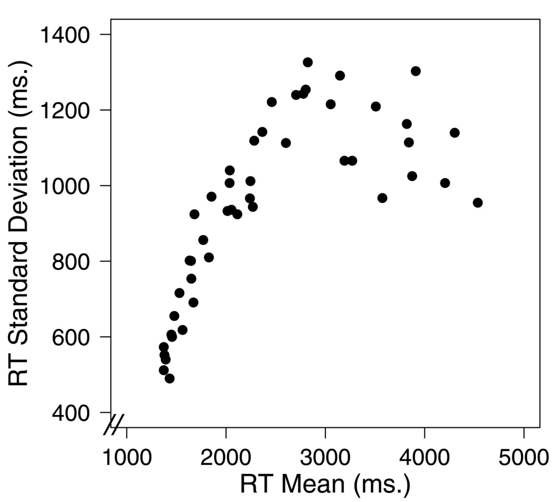


*Figure 5.* The nonlinear relation between response time (RT) mean and RT standard deviation (both in milliseconds) in Rickard's (2004) data. Mean and standard deviation were first computed over items for each participant and then averaged over participants.

## Theoretical and Practical Implications

The linear relation between RT mean and RT standard deviation as a function of task difficulty poses a constraint for the computational modeling of RT tasks. We have already shown that some descriptive RT distributions are inconsistent with the linear law (i.e., the shifted Wald distribution does not accommodate the linear law, at least not through a change in a single parameter). The linear constraint may also be of value for certain process models of RT. Consider, for instance, the multiple read-out model for lexical decision (Grainger & Jacobs, 1996). In lexical decision, task difficulty is mediated in part by word frequency (cf. Ratcliff et al., 2004; Wagenmakers, Ratcliff, et al., in press, for diffusion model applications). Thus, the linear law predicts that as word frequency increases, RT mean and RT standard deviation should decrease at the same rate. It is presently unclear whether the multiple read-out

[3] We thank Tim Rickard for sending us the data shown in Figure 5.

model produces this result. A similar implication holds for absolute identification models. Our detailed analyses of Brown et al.'s (2006) absolute identification data indicated that the relationship between mean RT and standard deviation RT was strongly linear in that paradigm. This linear relationship held over a wide variety of conditions, including changes in stimulus magnitude, changes in the size of the comparison set, and changes in task instructions given to participants. This result represents a challenge for models of absolute identification: It is unclear whether any of the current models can accommodate the linear relationship.

The linear law also corroborates a number of analysis techniques that implicitly assume a linear relation between RT mean and RT standard deviation. For instance, researchers in the field of aging are sometimes interested in the effects of variability after controlling for differences in mean performance (e.g., Hultsch et al., 2002; Li, 2002; MacDonald et al., 2003; Shammi et al., 1998). One method is to use a linear regression technique to partial out effects of differences in RT mean on the observed differences in RT standard deviation. Another method to control for baseline differences in processing speed is to use the coefficient of variation (i.e., standard deviation divided by the mean; e.g., Segalowitz & Segalowitz, 1993). These methods are only appropriate when the relation between RT mean and RT standard deviation is linear.

In psycholinguistics, the coefficient of variation has also been used to assess whether a practice-induced increase in language proficiency and lexical access is due to a simple speedup or a restructuring of the cognitive processes involved (e.g., Segalowitz & Frenkiel-Fishman, 2005; Segalowitz & Segalowitz, 1993). According to Segalowitz and colleagues (Segalowitz & Frenkiel-Fishman, 2005; Segalowitz & Segalowitz, 1993), a simple speedup is associated with a constant coefficient of variation (i.e., mean and standard deviation decrease at the same rate). When the coefficient of variation differs between participants or practice sessions, this is supposedly indicative of a shift toward a cognitively more efficient mode of processing. This idea is similar to that of Rickard's (1997) CMPL theory. Again, the linear law is consistent with this line of reasoning.

## Concluding Comments

We have shown that RT standard deviation is close to linear in mean RT for each individual participant, across a range of experimental tasks in memory, perception, categorization, and problem solving. The generality of this result is surprisingly strong: Nearly three quarters of participants had a correlation between mean and standard deviation of greater than .85. This strong empirical regularity can be considered a law of RT, like the law of practice (see Heathcote et al., 2000). Such laws can inform theory development—for example, Wagenmakers et al.'s (2005) results combine with our current results to provide support for Ratcliff's (1978) Wiener diffusion model and for the use of the random walk decision rule in Logan's ITAM model for attention, memory, and categorization (Logan, 2002). Other models of RT should also be analyzed to see whether they support the observed linear relationship between RT mean and standard deviation.

## References


Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review, 113,* 700–765.

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review, 112,* 117–128.

Brown, S. D., Marley, A. A. J., Donkin, C., & Heathcote, A. J. (2006). *An integrated, principled account of absolute identification*. Manuscript submitted for publication.

Carrasco, M., Ponte, D., Rechea, C., & Sampedro, M. J. (1998). "Transient structures:" The effects of practice and distractor grouping on within-dimension conjunction searches. *Perception & Psychophysics, 60,* 1243–1258.

Chessa, A. G., & Murre, J. M. J. (2006). Modelling memory processes and Internet response times: Weibull or power-law? *Physica A, 366,* 539–551.

Chocholle, R. (1940). Variations des temps de réaction auditifs en fonction de l'intensité a diverses fréquences [Variation of auditory response times as a function of the intensity at various frequencies]. *L'Année Psychologique, 41,* 65–124.

Christie, L. S., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *Bulletin of Mathematical Biophysics, 18,* 89–112.

Colonius, H. (1995). The instance theory of automaticity: Why the Weibull? *Psychological Review, 102,* 744–750.

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory based theories of automaticity. *Memory & Cognition, 19,* 151–158.

Cousineau, D., Goodman, V. W., & Shiffrin, R. M. (2002). Extending statistics of extremes to distributions varying on position and scale, and implications for race models. *Journal of Mathematical Psychology, 46,* 431–454.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9,* 1–7.

Fitts, P. M. (1992). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology: General, 121,* 262–269. (Reprinted from the *Journal of Experimental Psychology, 47,* 381–391)

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565.

Heathcote, A., Brown, S., & Cousineau, D. (2004). Estimating lognormal, Wald and Weibull RT distributions with a parameter dependent lower bound. *Behavior Research Methods, Instruments, & Computers, 36,* 277–290.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7,* 185–207.

Heathcote, A., & Mewhort, D. J. K. (1993). Selection and representation of relative position. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 448–515.

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Arnold, L. E., Conners, C. K., et al. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology, 12,* 125–140.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11–26.

Hohle, R. H. (1965). Inferred components of reaction time as a function of foreperiod duration. *Journal of Experimental Psychology, 69,* 382–386.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological Sciences, 57B,* 101–115.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45,* 188–196.

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

Lassaline, M. E., & Logan, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 561–581.

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

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

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22,* 1–35.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 883–914.

Logan, G. D. (1995). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review, 102,* 751–756.

Logan, G. D. (1998). What is learned during automatization? II: Obligatory encoding of spatial location. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1720–1736.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109,* 376–400.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–195.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

MacDonald, S. W. S., Hultsch, D. F., &. Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria longitudinal study. *Psychology and Aging, 18,* 510–523.

McMillen, T., & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology, 50,* 30–57.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354.

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

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

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

Ratcliff, R., & 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.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126,* 288–311.

Rickard, T. C. (1999). A CMPL alternative account of practice effects in numerosity judgment tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 532–542.

Rickard, T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 65–82.

Rickard, T. C., & Bourne, L. E., Jr. (1996). Some tests of an identical elements model of basic arithmetic skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1281–1295.

Rouder, J. N., Lu, J., Speckman, P. L., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review, 12,* 195–223.

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition, 33,* 644–653.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14,* 369–385.

Shammi, P., Bosman, E., & Stuss, D. T. (1998). Aging and variability in performance. *Aging, Neuropsychology, and Cognition, 5,* 1–13.

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biology, 3,* 334–349.

Sigman, M., & Dehaene, S. (2006). Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLoS Biology, 4,* e220.

Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 506–526.

Sternberg, S. (1966, August 5). High-speed scanning in human memory. *Science, 153,* 652–654.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* London: Cambridge University Press.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592.

Wagenmakers, E.-J., Grasman, R. P. P. P., & Molenaar, P. C. M. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology, 49,* 195–204.

Wagenmakers, E.-J., & Grünwald, P. (2006). A Bayesian perspective on hypothesis testing. *Psychological Science, 17,* 641–642.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (in press). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language.*

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

Wagenmakers, E.-J., & Waldorp, L. (Eds.). (2006). Model selection: Theoretical developments and applications [Special issue]. *Journal of Mathematical Psychology, 50*(2).

(*Appendixes follow*)

# Appendix A

## Details of the Descriptive Response Time (RT) Distributions

Take the most commonly used descriptive distributions for RT analysis: the ex-Gaussian, the shifted lognormal, the shifted Wald, the shifted Weibull, and the Gumbel (e.g., Heathcote, Brown, & Cousineau, 2004). For each of these distributions, we consider the relation between mean and standard deviation as a function of its parameters.

### *Ex-Gaussian*

The three parameters of the ex-Gaussian distribution are the mean ($\mu$) and standard deviation ($\sigma$) of the normal component and the mean ($\tau$) of the exponential component. The mean of the ex-Gaussian distribution is $\mu + \tau$, and its standard deviation is $\sqrt{\sigma^2 + \tau^2}$, so as long as $\tau >> \sigma$, changes in $\tau$ will lead to a linear relationship between mean and standard deviation. This analysis suggests that $\tau$ should play the role of representing task difficulty in the ex-Gaussian distribution, a conclusion that prior analyses have supported (cf. Spieler, Balota, & Faust, 2000, p. 519). The use of $\tau$ to represent task difficulty also agrees with Hohle's (1965) original motivation for developing the ex-Gaussian distribution, in which the exponential component (indexed by $\tau$) was intended to represent the decision processing time (cf. Christie & Luce, 1956).

### *Shifted Lognormal*

The shifted lognormal distribution has two parameters representing the mean of the underlying normal distribution (mean $\mu$ and standard deviation $\sigma$) as well as a shift parameter $\theta$. The mean of the shifted lognormal distribution is $\theta + \exp\left(\mu + \frac{1}{2}\sigma^2\right)$, and the standard deviation is given by $[\exp(2\mu + \sigma^2) \cdot (\exp(\sigma^2) - 1)]^{1/2}$. The logarithmic nature of the parameters in this distribution produces linear relationships: Changes in $\mu$ will always lead to linearity of the mean versus standard deviation. Changes in $\sigma$ will also lead to linearity, as long as $\sigma$ is larger than about 2, which ensures that $(\exp(\sigma^2) - 1) \sim \exp(\sigma^2)$ to a reasonable approximation.

### *Shifted Wald*

The shifted Wald distribution represents the density of first passage times of a drifting Brownian motion process through a single absorbing barrier, along with a positive offset. The parameters of the distribution are the drift rate of the process ($\mu$), the separation between the starting point of the motion and the barrier ($a$), and the shift value ($\theta$). The mean of the resulting distribution is given by $\theta + a / \mu$ and the standard deviation is given by $\sqrt{a/\mu^3}$. If one treats the $\mu$ as indexing task difficulty, then the relationship between mean and standard deviation is not linear: The standard deviation is then close to linear in the mean raised to the power of 1.5. If one treats $a$ as indexing task difficulty, then again the relationship is nonlinear: The standard deviation is linear in the square root of the mean RT.

### *Shifted Weibull*

The shifted Weibull distribution is a power transformation of an exponential distribution and arises as one of the three extreme value distributions (for applications to psychology, see Chessa & Murre, 2006; Logan, 1988, 1992; see also Cousineau et al., 2002). This distribution has a shift parameter ($\theta$), a parameter for the mean of the underlying exponential distribution ($\tau$), and a power transformation parameter ($c$). The mean of the shifted Weibull distribution is given by $\theta + \tau\Gamma(c^{-1}+1)$, and the standard deviation is given by $\tau[\Gamma(2c^{-1} + 1) - \Gamma^2(c^{-1} + 1)]^{1/2}$ , where $\Gamma$ represents the incomplete gamma function. If the mean of the underlying exponential distribution ($\tau$) represents task difficulty, then the relationship between mean and standard deviation is perfectly linear, neglecting the effects of the shift parameter ($\theta$).

### *Gumbel*

The Gumbel distribution is another extreme value distribution, with only two parameters: a location parameter ($\mu$) and a scale parameter ($\sigma$). The mean of the Gumbel distribution is given by $\mu + 0.578\sigma$, and the standard deviation is given by $\sigma\pi/\sqrt{6}$. Therefore, task difficulty must be indexed by the $\sigma$ parameter for a perfectly linear relationship between mean and standard deviation.

# Appendix B

## Description of Experiments

In addition to the absolute identification experiment, we examined nine sets of data; collectively, the data represent 2,452 experimental conditions from 127 participants. Each data set was given an acronym used to index the summaries of results (see Table 2). The following sections describe both the paradigms from which the data were drawn and the experimental factors used to produce separate series for each data set. The data sets are freely available online at http://www.newcastle.edu.au/school/psychology/ncl/data_repository.html.

### *Counting*

In the counting tasks, participants were shown different patterns of 6 to 11 dots and a spelled-out number; they were asked verify whether the number of dots in the pattern matched the spelled-out number. All data were taken from Palmeri (1997). Each experiment used a number of unique patterns, and fits included series from each pattern. Fits used data broken down by participants and dot pattern. The data in set C1 are the training series from Experiment 1. The number of dots was manipulated within-subjects. There were 30 patterns with 5 patterns of array size. The data in set C2 are the training series from Experiment 2. The number of dots and the similarity of dot patterns were manipulated within-subjects. There were 72 patterns with 4 patterns for each level of similarity per array size. The data in set C3 are the training series from Experiment 3. The number of the dots and similarity were manipulated within-subjects. There were 72 patterns with 6 patterns for each level of similarity at each array size.

### *Mental Arithmetic*

The mental arithmetic tasks included a diverse set of problem types. The data in set M1 are from a single-digit multiplication task taken from Experiment 1 of Rickard and Bourne (1996). Either the participants were shown two digits and asked to calculate the product or they were shown a digit and a product and asked to divide the product to compute the dividend. Response time (RT) was recorded as the time between the presentation of the problem and the keystroke of the first digit of the answer. There were 16 problem examples. Problem type (compute product or compute dividend) and range of digits was manipulated within-subjects. The data in set M2 are from a three-step arithmetic task (Experiment 1 of Rickard, 1997). Participants were shown two numbers and asked to calculate their difference, to add 1 to the result, and, then, replacing one of the numbers with the result so far, to compute the sum of it with the remaining original number. RT was recorded as the time between the presentation of the problem and the keystroke of the first digit of the answer.

### *Alphabetic Arithmetic*

The data in set A1 are from Experiment 2 of Rickard (1997), in which participants were required to verify equations of the form $A + 2 = C$ or $A + 3 = C$ (true and false equations), respectively. We broke down the data by participants and by problem example. Two factors were manipulated within-subjects: addend (3, 5, and 7) and trial type (true–false), with four examples of each type.

### *Visual Search*

In the visual search tasks, participants were required to indicate whether a target appeared in a visual display. In V1 and V3, the target was defined by the relative position of two features; in V2, the target was defined by a conjunction of colors. Stimuli used for targets and distractors were consistently mapped over trials in V1 and V2. Targets and distractors were variably mapped in V3, and a target cue was given before each trial. Data in set V1 are from Experiment 1 of Heathcote and Mewhort (1993). Two factors were manipulated between-subjects: feature type (brightness or color) and display area (small or large). Two factors were manipulated within-subjects: display size (two, four, six, or eight objects) and trial type (target–distractor). The data in set V2 are from Experiment 3 of Carrasco, Ponte, Rechea, and Sampedro (1998). Two factors were manipulated within-subject: display size (2, 6 or 10, 14, 18, 22 objects) and trial type (target–distractor). The data in set V3 are from Experiment 3 of Heathcote and Mewhort (1993). Three factors were manipulated within-subject: display size (two, four, six, or eight objects), target type, and trial type (target–distractor).

Received October 6, 2005
Revision received December 5, 2006
Accepted January 25, 2007 ■