ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
000(00) 1-7

DOI: 10.1177/0963721411408885
http://cdps.sagepub.com
SAGE

Moving Beyond the Mean in Studies of Mental Chronometry: The Power of Response Time Distributional Analyses

David A. Balota[1] and Melvin J. Yap[2]
[1] Washington University in St. Louis and [2] National University of Singapore

Abstract
Although it is widely recognized that response time (RT) distributions are almost always positively skewed and that mathematical psychologists have developed straightforward procedures for capturing characteristics of RT distributions, researchers continue to rely primarily on mean performance, which can be misleading for such data. We review simple procedures for capturing characteristics of underlying RT distributions and show how such procedures have recently been useful to better understand effects from standard cognitive experimental paradigms and individual differences in performance. These well-studied procedures for understanding RT distributions indicate that effects in means can be produced by (a) shifts of RT distributions, (b) stretching of slow tails of RT distributions, or (c) some combination. Importantly, effects in means can actually be obscured by opposing influences on the modal and tail portions of RT distributions. Such disparate patterns demand novel theoretical interpretations.

Keywords
response time distributional analyses, ex-Gaussian analyses, visual word recognition, individual differences

We all love the mean! The mean naturally provides an easy summary statistic for multiple observations, and as a result, means are common in our everyday lives (e.g., average miles per gallon, batting average, political approval rating). However, anyone who has taken an introductory statistics course knows that means are only one way to provide an estimate for a distribution of numbers and that they really work best for distributions that are relatively symmetrical in shape around some mode. Importantly, means can be misleading. For example, consider a politician's mean approval rating. The mean may only represent a minority of voters since approval rating is likely to reflect a bimodal distribution based on party allegiance. Distributions can also be asymmetrical around a single mode. Empirical response time (RT) distributions, the focus of the present article, are virtually always positively skewed, with RTs clustering at the faster end of the scale. Given our knowledge of the shape of RT distributions, it is surprising that the vast majority of studies of mental chronometry still rely on mean RT performance. For example, we recently reviewed 285 articles published in 2010 in three leading experimental journals, and found that 49% of the studies used RT measures. Importantly, of those articles that used RT as a dependent measure, 95% relied primarily on mean RT.

We first briefly describe alternative methods for capturing the effects of manipulations on the underlying distributions in RT experiments. This is followed by a discussion of what we have recently learned from a series of studies from our laboratory that have examined RT distributions. These studies illustrate the distinct patterns one can obtain when examining underlying RT distributions. It is important to emphasize that these are not new observations (e.g., Heathcote, Popiel, & Mewhort, 1991; Luce, 1986; Ratcliff, 1979; Ratcliff & Murdock, 1976). Our goal here is simply to build on this earlier pioneering work to further demonstrate the advantages of examining the effects of standard cognitive manipulations on RT distributions.

There are three general approaches for understanding the influences of variables on RT distributions. The first and ultimately the preferred way is to use a computationally explicit model that makes specific predictions about the characteristics of RT distributions. The most common example here is the diffusion model (Ratcliff, 1978), which makes predictions about the RT distributions for both correct and incorrect responses in binary-decision tasks. Second, one can evaluate the influence of manipulations on the parameters derived from a

Corresponding Author:
David A. Balota, Department of Psychology, Washington University, St. Louis, MO 63130
E-mail: dbalota@artsci.wustl.edu

mathematical function that has been fit to an empirically obtained RT distribution. The third approach is to simply plot the shape of the RT distribution to determine how a manipulation changes the different regions of the distribution. Because many RT studies do not rely on predictions from explicit computational models, we focus here primarily on the latter two approaches.[1]

Fitting a Mathematical Function to an RT Distribution

A useful approach to understanding the influence of a manipulation is to fit a mathematical function to an RT distribution to assess how different parameters of the function are modulated by experimental manipulations. There are many useful and well-studied functions, including the ex-Gaussian, Wald, Gamma, and Weibull, among many others (see Van Zandt, 2000, for a review of the advantages and disadvantages of each function). Here we focus on the ex-Gaussian function, which is the convolution (a mathematical combination) of a Gaussian and an exponential distribution. The mode and standard deviation of the Gaussian component are approximated by μ and σ respectively, while the exponential function is approximated by τ, which reflects the mean and standard deviation of the exponential component.

The ex-Gaussian function possesses a number of useful properties, as Ratcliff and Murdock (1976) and Heathcote et al. (1991) have noted. First, the function provides excellent fits for empirically obtained RT distributions (see also Luce, 1986). Second, and importantly, parameters from the ex-Gaussian function algebraically map onto the mean of an RT distribution. Specifically, the mean of an empirically obtained RT distribution is constrained to be the sum of μ and τ. Hence, because of this constraint, the ex-Gaussian approach takes an important step toward making contact with the mean-dominated literature.[2]

Figure 1 displays how a variable might influence the characteristics of an RT distribution as reflected by ex-Gaussian parameters. Comparing the top and second panel, one sees a simple shift in the RT distribution, as reflected by a change only in μ. Comparing the top and third panel, one can see only a stretching of the tail of the distribution, as reflected by a change in τ. Comparing panels A and D, one can see a tradeoff between μ and τ, such that there are two counteracting influences at the distributional level but no difference in the means. Importantly, all three patterns have now been well established in the experimental literature.

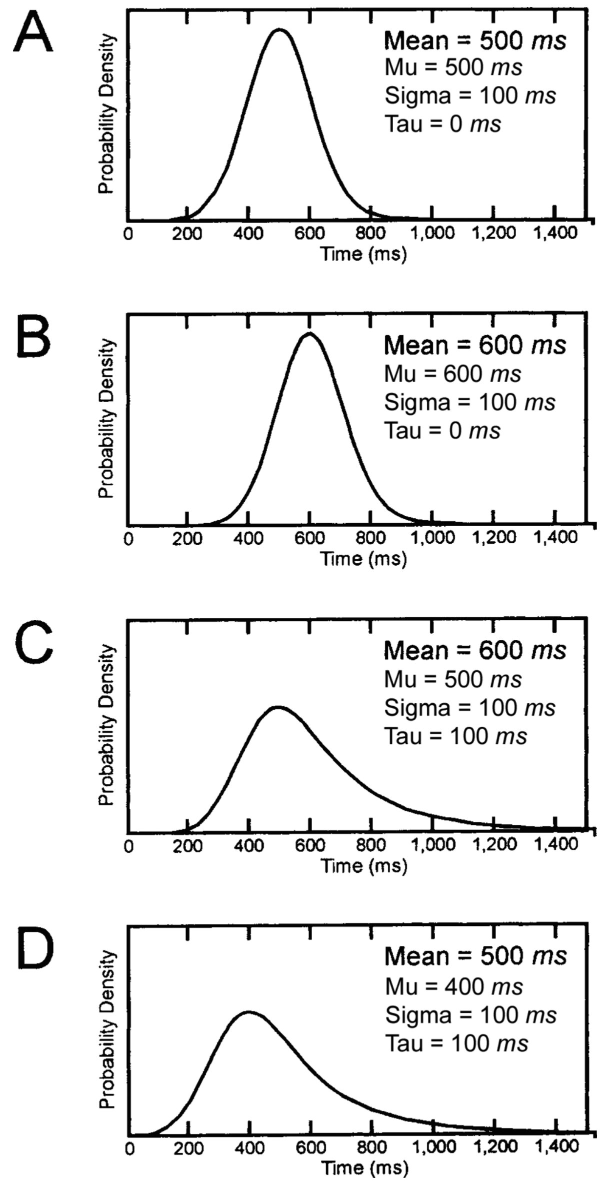

Fig. 1. Possible changes in distributions and the underlying influences on mean estimates and the parameter estimates from ex-Gaussian analyses. Adapted from "Word Frequency Repetition, and Lexicality Effects in Word Recognition Tasks: Beyond Measures of Central Tendency," by D.A. Balota and D.H. Spieler, 1999, *Journal of Experimental Psychology: General*, *128*, p. 33. Copyright 1999, American Psychology Association.

Descriptive Plots of RT Distributions

Just as there are multiple mathematical functions to fit an empirically obtained RT distribution, there are also multiple ways to plot RT distributions. These include Hazard, Delta, Quantile, and Vincentile plots. Simply plotting the data is very useful whenever one considers effects on underlying distributions, and there should be convergence between such plots and parameter estimates obtained by fitting empirical data to a mathematical function. Here, for simplicity, we describe Quantile plots. In quantile[3] analyses, one rank orders the RTs for a given participant as a function of condition and plots the quantiles (.1, .2, .3, etc.). One can then plot the difference in the quantiles to better understand how a variable influences different portions of the underlying RT distribution. A related procedure involves Delta plots, which display the effects of variables as a function of mean response latency (see De Jong, Liang, & Lauber, 1994).

Convergence Between Ex-Gaussian Parameters and Quantiles

It is important to understand the relation between function fitting and plotting the effects of a variable on empirical RT distributions. In order to address this, we used Monte Carlo simulations to generate ex-Gaussian distributions (comprising 10,000 observations for each distribution) with varying levels of each parameter (see Fig. 2 legends for specific parameter values used). For example, the top two panels of Fig. 2 display the quantiles for two hypothetical conditions. The top panel displays an effect on μ, whereas the middle panel displays an effect on τ. The bottom panel displays the plot of differences across the quantiles. As shown, a change in μ produces a simple shift across the quantiles, whereas a change in τ produces a larger effect in the slower quantiles.[4]

What Have We Learned Beyond the Mean?

Effects of individual factors

For illustrative purposes, we will focus on some of our own work (since quite distinct patterns have been observed) on the influence of standard cognitive manipulations on RT distributions. First, consider the classic semantic-priming effect in the lexical-decision (discriminating between real words and nonwords, e.g., FLIRP) and the speeded-pronunciation tasks (reading words aloud). Participants recognize a word (e.g., CAT) faster when it is preceded by a related word (DOG–CAT) than by an unrelated word (DIG–CAT). The results of a recent series of experiments indicate that for highly skilled readers, semantic-priming effects are purely reflected in distributional shifting—that is, priming modulates μ, but not σ or τ (Balota, Yap, Cortese, & Watson, 2008). This is consistent with a simple head-start model of priming whereby the related prime produces a constant benefit that is independent of target difficulty. This contrasts well with word-frequency effects in lexical-decision performance, whereby word frequency shifts *and* increases the tail of the RT distribution (e.g., Balota & Spieler, 1999), reflecting a larger influence of frequency on the more difficult items (see priming and frequency effects in the top panel of Fig. 3).

Interestingly, there is also clear evidence of trade-offs between parameters. Consider the classic Stroop color naming effect. Heathcote et al. (1991) and Spieler, Balota, and Faust (1996) found that the inconsistent *facilitation* in mean Stroop color naming RT performance (i.e., naming the color of the word RED printed in red compared to naming the color of the word DEEP printed in red) was due to facilitation of the modal portion in the congruent condition, which was offset by an increase in the tail of the distribution, thereby canceling each other out in the means.

Studies have also shown that different tasks that presumably reflect similar mechanisms produce different influences on the underlying RT distributions. For example, the *congruency effect* (incongruent vs. congruent) in the Stroop task and the Simon task (a spatial-selection task) are often viewed as reflecting similar attentional selection mechanisms. However, Castel, Balota, Hutchison, Logan, and Yap (2007) have shown that the congruency effect in the Simon task is largest for the fastest responses, whereas Spieler et al. (1996) have shown that the congruency effect in the Stroop task is larger in the tail of the RT distribution (see also Pratte, Rouder, Morey, & Feng, 2010).

Joint effects of factors

Researchers are most often interested in the combined effects of multiple independent variables. For example, there is a large literature indicating that semantic priming interacts with target degradation, such that visually degraded targets show more priming than clear targets do. As shown in the middle panel of Figure 3, this interaction is primarily driven by the slow tail of the RT distribution (see Balota et al., 2008).

In contrast to these interactive effects, word frequency and stimulus degradation produce additive effects in lexical-decision performance. These additive effects occur in all parameters of the ex-Gaussian function and are also seen in Quantile analyses (e.g., Yap, Balota, Tse, & Besner, 2008). Interestingly, however, when one makes the discrimination between words and nonwords more difficult in lexical decision by using pseudohomophones (nonwords that sound like real words, e.g., BRANE), compared to more typical pronounceable nonwords (e.g., FLIRP), one still finds additive effects of word frequency and stimulus degradation in the means but interactive opposing effects in the modal portion and the tail of the RT distributions, respectively reflected in μ and τ. This latter intriguing pattern has been replicated in three different universities (see Yap et al., 2008), supporting the robustness of these distributional analyses (see bottom panel of Fig. 3).

Insights Into Individual Differences

An important central issue in cognitive science is the nature of individual differences in cognitive components. Regarding RT distributions, it does appear that individuals carry with them their own characteristic RT distributions that are relatively stable over time. For example, Yap, Balota, Sibley, and Ratcliff (2010) recently investigated the test–retest correlations across days of testing and found that the correlations for μ, σ, and τ were .717, .509, and .872 for lexical decision, and .865, .732, and .940 for pronunciation, respectively, even though different stimuli were presented across the days. It is noteworthy that the that the tail (τ) of the RT distribution appears to be more stable than the modal (μ) portion of the distribution, and indeed the test–retest reliability for τ was comparable to the correlations in the means (.871 and .929, in lexical decision and pronunciation, respectively).

One of the most studied aspects of individual difference in cognition is working-memory capacity (see Engle, Tuholski, Laughlin, & Conway, 1999). Schmiedek, Oberauer, Wilhelm, Süß, and Wittmann (2007) explored the relationship between working-memory capacity and the ex-Gaussian parameters

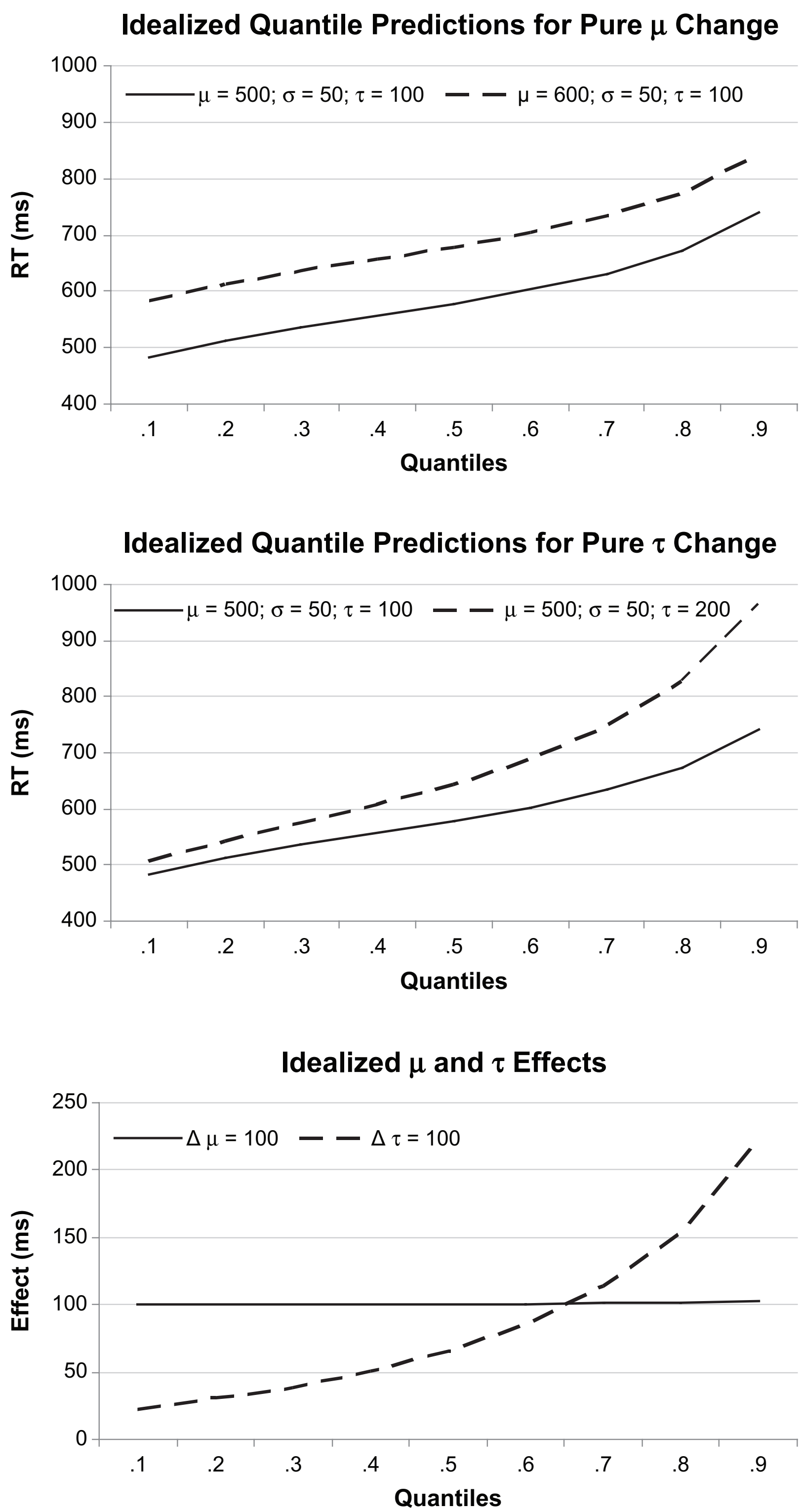

Fig. 2. Isolated effects of changes in the ex-Gaussian parameters on the underlying quantiles based on Monte Carlo simulations, where each distribution comprised 10,000 observations. The top panel displays an effect on μ, whereas the middle panel displays an effect on τ. The bottom panel displays the plot of difference scores across conditions taken from the top two panels. RT = reaction time.

estimated from an independent set of RT tasks. The results were surprisingly clear: τ was by far the parameter most strongly related to working memory. Tse, Balota, Yap, Duchek, and McCabe (2010), using similar analyses, have replicated the same pattern in a group of older adults using different RT tasks and working-memory measures. Specifically, the path between

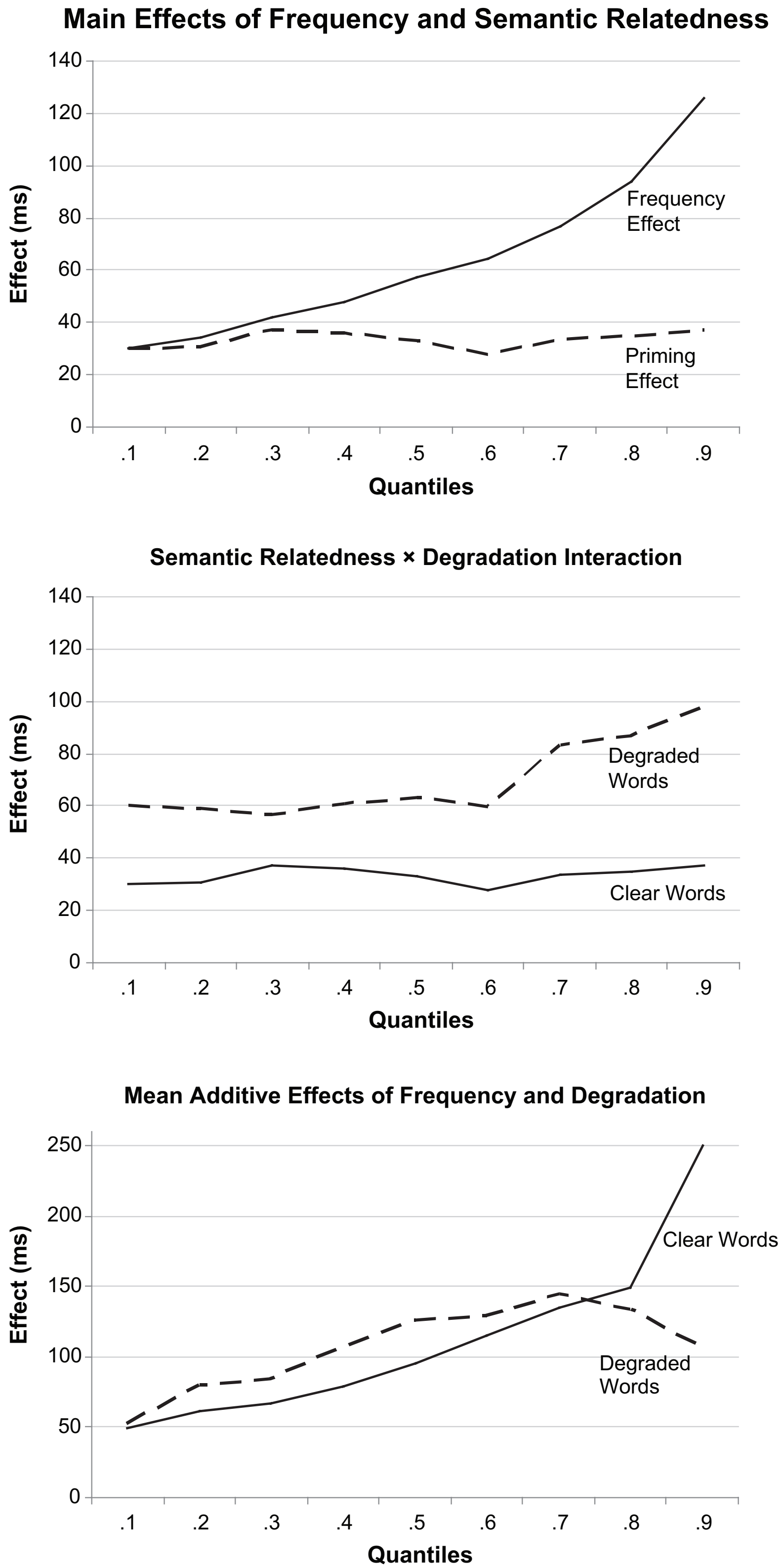

Fig. 3. Main effects of semantic relatedness and word frequency across quantiles, showing that semantic priming shifts reaction time distributions, whereas word frequency both shifts and stretches the tail of the distribution (top panel); interactive effects of semantic priming and visual degradation, showing that the larger semantic priming effects for degraded words than clear words is due to both a shift and stretching of the tail of the distribution (middle panel); and a pattern in which the additive effects of word frequency and stimulus degradation observed in the means reflect tradeoffs between the modal portion and tail of the distribution (bottom panel).

τ and working-memory construct was –.90, whereas the paths for μ and σ were .24 and –.09, respectively.

The robust relationship between τ and working memory is intriguing in light of the worst-performance rule developed in the intelligence literature (e.g., Coyle, 2003). According to this principle, participants' slow RTs are more strongly related to fluid intelligence than are their fast RTs. Given that working-memory performance is strongly related to fluid intelligence and τ is related to the positive skew of RT distributions, the relationship between τ and working memory is consistent with the worst-performance rule. This link between performance on the most difficult trials and fluid intelligence further underscores the importance of examining the slow tail of RT distributions.

Distinguishing Between Parameters and Processes

Although tempting, we need to emphasize that one should be careful not to simply map processes onto parameters or aspects of RT distributions without converging evidence (Matzke & Wagenmakers, 2009). For example, as Schmiedek et al. (2007) point out, the relationship between working memory and τ observed in their data is also compatible with changes in drift rate, a parameter in Ratcliff's (1978) diffusion model (mentioned earlier) that reflects how efficiently people accumulate information about a stimulus. Drift-rate changes do not demand inferences about working memory. Thus, at this point, it is important to appreciate the limitations of mapping process onto parameters. Just as in developing an understanding of the mechanistic influence of a manipulation on mean RTs, one also needs to rely on converging evidence across studies, in tandem with explicit modeling, to constrain interpretations of how variables influence the underlying RT distribution.

Conclusions

The goal of the present article is to further encourage researchers to look beyond measures of central tendency to better understand the influence of manipulations on performance. Following early pioneering work (e.g., Ratcliff & Murdock, 1976; Heathcote et al., 1991), we have provided further evidence that fitting an empirically obtained RT distribution to a mathematical function (e.g., ex-Gaussian) and simply plotting the RT distributions as a function of condition afford significant advances over analyses of means in standard experimental paradigms. In exemplifying this approach, we have shown how distributional analyses can provide insights into individual differences across the lifespan and into commonly used experimental paradigms (e.g., word-recognition, semantic-priming, Simon, and Stroop tasks). Advances in science are often made in step with an increase in the sensitivity of the measuring device, and we continue the appeal that it is time to increase the sensitivity of the microscope used to study mental chronometry.

Recommended Reading

Heathcote, A., Popiel, S.J., & Mewhort, D.J.K. (1991). (See References). The first article to show how ex-Gaussian parameters tradeoff to minimize effects in the Stroop task.

Luce (1986). (See References). Still the most comprehensive book on RT distribution modeling.

Ratcliff, R. (1978). (See References). An excellent paper showing how RT distributions change as a function of drift rate in the diffusion model.

Ratcliff, R. (in press). Response time distributions. In H. Cooper (Ed.), Handbook of research methods in psychology. Washington, DC: American Psychological Association. A recent overview of progress in understanding response time distributions.

Van Zandt, T. (2000). (See References). A detailed review of the advantages and disadvantages of different RT fitting functions.

Acknowledgments

We thank Andrew Heathcote, James Neely, Roger Ratcliff, Keith Rayner, Jeffery Rouder, and E.J. Wagenmakers for helpful comments regarding this work.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The work in this article was supported by National Institute on Aging AG03991 and National Science Foundation BCS 0001801 grants to D.A. Balota.

Notes

1. Free and simple-to-use software packages are readily available that enable any user to carry out distributional analyses on their own RT data. For example, QMPE 2.18 (http://www.newcl.org/software/qmpe.htm; Brown & Heathcote, 2003) affords easy fitting of RT data with as few as 40 observations per condition.
2. Ex-Gaussian parameters can also be used to approximate the variance ($\sigma 2 + \tau 2$) and skew of a distribution ($2\tau 3$).
3. Quantiles are conceptually similar to percentiles, except that the former range from 0 to 1 while the latter range from 0 to 100. Instead of plotting quantiles, one can also plot Vincentiles, which reflect the mean of a group of scores for a given individual (e.g., 1 to 10%, 11 to 20%, etc.) averaged across participants as a function of condition (Gilchrist, 2000). In practice, the two methods yield very similar results (Jiang, Rouder, & Speckman, 2004).
4. For simplicity, we have not shown how changes in σ are reflected in underlying RT distributions, but see Balota, Yap, Cortese, and Watson (2008) for a discussion. Of course, these are idealized changes in single parameters with the goal of making contact with the mean-dominated RT literature.

References

Balota, D.A., & Spieler, D.H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*, 32–55.

Balota, D.A., Yap, M.J., Cortese, M.J., & Watson, J.M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory & Language*, *59*, 495–523.

Brown, S., & Heathcote, A. (2003). QMLE: Fast, robust and efficient estimation of distribution functions based on quantiles. *Behavior Research Methods*, *Instruments & Computers, 35*, 485–492.

Castel, A.D., Balota, D.A., Hutchison, K.A., Logan, J.M., & Yap, M.J. (2007). Spatial attention and response control in healthy younger and older adults and individuals with Alzheimer's disease: Evidence for disproportionate selection breakdowns in the Simon task. *Neuropsychology*, *21*, 170–182.

Coyle, T.R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence*, *31*, 567–587.

De Jong, R., Liang, C.C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus–response concordance. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 731–750.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Gilchrist, W.G. (2000). *Statistical modelling with quantile functions*. Boca Raton, FL: Chapman & Hall.

Heathcote, A., Popiel, S.J., & Mewhort, D.J.K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*, 340–347.

Jiang, Y., Rouder, J.N., & Speckman, P.L. (2004). A note on the sampling properties of the Vincentizing (quantile averaging) procedure. *Journal of Mathematical Psychology*, *48*, 186–195.

Luce, R.D. (1986). *Response times: Their role in inferring elementary mental organization*.

Matzke, D. & Wagenmakers, E.-J. (2009). Psychological interpretation of ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*, 798–817.

Pratte, M.S., Rouder, J.N., Morey, R.D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception & Psychophysics*, *72*, 2013–2025.

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

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

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

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W.W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, 414–429.

Spieler, D.H., Balota, D.A., & Faust, M.E. (1996). Stroop performance in normal older adults and individuals with senile dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 461–479.

Tse, C.S., Balota, D.A., Yap, M.J. Duchek, J.M., & McCabe, D.P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology*, *24*, 300–315.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*, 424–465.

Yap, M.J., Balota, D.A., Sibley, D., Ratcliff, R. (2010, November). *Individual differences in visual word recognition: Insights from the ELP*. Poster presented at the 51st annual meeting of the Psychonomic Society, St. Louis, MO.

Yap, M.J., Balota, D.A., Tse, C.S., & Besner, D. (2008). On the interactive effects of stimulus quality and word frequency in lexical decisions: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *34*, 495–513.