An Instrumental Cognitive Model for Speeded and/or Simple Response Tasks

Royce Anders (royce.anders@univ-amu.fr)
LPC UMR 7290 CNRS, Aix-Marseille Université
3 place Victor Hugo, 13331, Marseille Cedex 3, France

F.-Xavier Alario (francois-xavier.alario@univ-amu.fr)
LPC UMR 7290 CNRS, Aix-Marseille Université
3 place Victor Hugo, 13331, Marseille Cedex 3, France

Leendert van Maanen (l.vanmaanen@uva.nl)
Department of Psychology, University of Amsterdam
Weesperplein 4, 1018 XA, Amsterdam, The Netherlands

Abstract

Speeded and/or simple response tasks may be cognitively modeled by a random walk process that accumulates to threshold. In cases of tasks where mainly one characteristic response is observed, at varying latencies, then random walks involving only positive drifts that each arrive at a single threshold, provide a suitable accumulation modeling account of the data; and advantageously, this accumulation model is exactly described by the shifted Wald (SW) probability density function. We will demonstrate how the SW distribution is thus a noteworthy cognitive model for these tasks, which uniquely possesses simultaneously, high utility as an objective data measurement tool for the response time (RT) distributions. Per each experiment condition, its three parameters can decompose the observed mean RT value, quantify the shape and characteristics of the observed RT distribution, and account for significant differences between distributions with near-identical mean values; regardless of whether one accepts the cognitive interpretation of the random-walk accumulation process. We present the SW model and demonstrate its efficiency and utility on both simulated and real data.

Keywords: response time analysis, shifted Wald, psychometrics, accumulation modeling

Introduction

In the psychological sciences, the efficacy of modeling the distributions of response time (RT) data, rather than only using classical methods, to obtain a deeper understanding of experiment effects and underlying processes, has been well-demonstrated in the preceding literature (Ratcliff, 1978; Luce, 1986; Ratcliff & Rouder, 1998; Andrews & Heathcote, 2001; Heathcote, 2004; Van Zandt, 2000, 2002; Ratcliff et al., 2004; Balota et al., 2008; Van Maanen et al., 2012; Staub et al., 2010; Balota & Yap, 2011). In the present paper we bring attention to a simple-yet-powerful tool for RT data analysis, that despite its utility, is not yet in general use within the psychological community.

There exist quantitative distribution measurement tools for RT data, in which the parameters describe the properties of the observed data distribution; these tools are typically closed-form probability density functions with positive skew and values, such as the shifted Wald (SW, see Chapter 8.2 Luce, 1986; Heathcote, 2004), ex-Gaussian (Heathcote et al., 1991), shifted Weibull, shifted log-normal, and Gumbel (Wagenmakers & Brown, 2007). Then there are more complicated models of RT data that model signal accumulation: such as the Linear Ballistic Accumulator (LBA, Brown & Heathcote, 2008), race model (LaBerge, 1962), and the Drift Diffusion Model (DDM, Ratcliff & Murdock, 1976; Ratcliff, 1978; Ratcliff & McKoon, 2008), however their parameters do not directly describe the distribution of RT data. We bring to attention that uniquely, the SW distribution does both at the same time, and argue that as an accumulation model, it is on par in usefulness with more complex models of accumulation, when used in the appropriate context.

The Shifted Wald

Among a number of situations, the SW for RT data is apt for experiments consisting of speeded (e.g. 500ms-2000ms) and/or simple response tasks, where in particular, the ratio of errors to correct responses is small; some concrete examples consist of visual search, picture-naming, simple detection, and go no-go tasks. One should note that being only a distribution that is fit, the SW is a very simplistic model with few assumptions. However despite its simplicity, it possesses a formidable characteristic that stands it apart from the other distributions and models listed: while its parameters directly quantify the RT distribution, they also simultaneously, directly describe the RT values in the context of a Brownian motion process (BMP) in which a latent quantity accumulates to threshold; this is the same kind of BMP, related throughout the literature to signify the signal-to-response threshold event, that is at the root of the other popular signal-accumulation models, such as the DDM, race, and LBA models. Thus while being a quantitative measurement tool that can be applied to describe the distribution of any set of magnitude RTs, the SW model also provides an opportunity for theoretical work, such as on the cognitive-behavioral response process, based on its ability to also describe the data in terms of latent signal accumulation.

As an Accumulation Model

The SW with parameters, γ, α, and θ, can directly describe the data in the context of a continuous time-stochastic process (a type of BMP), consisting of a single latent quantity, X, that is continuously accumulating until it reaches a threshold. More specifically, X, accumulates at a given rate, γ, with

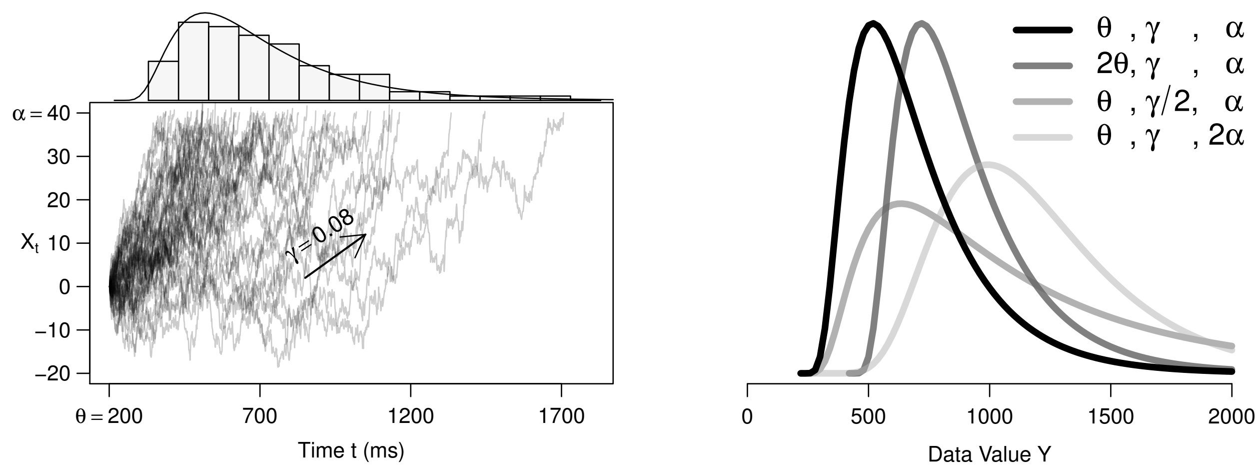

Figure 1: The SW as a cognitive-behavioral model (left), describing the RT data in the context of a latent quantity (e.g. signal) accumulating to threshold, α, at rate, γ, where θ accounts for the time lapsed outside of (around) this process. Then the SW as a distribution measurement tool (right). The black distribution has $\theta = 200$, $\gamma = 0.08$, and $\alpha = 40$, and illustrated in different shades of grey, are individual parameter adjustments that each cause unique distribution outcomes. Here they are each adjusted in the direction that results in bigger data values (e.g. slower RTs, slower mean RTs), which in each instance, results by a different distribution form.

noise until it reaches a threshold, α; and θ (the shift) is the minimal time lapsed outside of the process, which can be distributed before and after this accumulation process; the total time lapsed, T, is the data fit by the SW. This latent accumulation process provides a potential model for any data that involves a quantity accumulating over time that eventually reaches a value (or threshold). The SW thus provides the opportunity for a potentially-useful model, analogous to the signal-to-response threshold event of behavior.

In the context of RT data and the appropriate experimental task, this kind of underlying accumulation process that we note is similarly shared (by elementary adjustments) with the other aforementioned accumulation models, has been well-supported to correspond to a signal-to-response threshold, cognitive-behavioral event. In the case of the signal-to-response threshold interpretation of the SW: γ corresponds to the accumulation rate of the internal signal X, α to the threshold needed to initiate the physical response, and θ to the time distributed before and after this process (thus time lapsed outside of signal accumulation). The total time lapsed, T, is the RT recorded.

This latent accumulation process is illustrated in the left plot of Figure 1, in which many *random walks with drift* (RWDs, starting at $\theta = 200$, and having average slope $\gamma = 0.08$) as they intercept threshold $\alpha = 40$, are shown to correspond to a SW distribution with the same parameters: $\{\gamma = 0.08\,, \alpha = 40\,, \theta = 200\}$. Each of these RWDs are of the form

$$X_t = X_{t-1} + \gamma + \varepsilon\,, \tag{1}$$

where the position of a random variable X at time t, as X_t, is equal to its prior position value, X_{t-1}, plus a movement tendency, $\gamma > 0$ (known as drift), and marginal error, ε (or noise).[1]

Then note that any given threshold, $\alpha > 0$, unto which the time process terminates when X_t reaches that value, as $X_t \geq \alpha$, will produce a Wald distribution of data: letting T denote the time t at which X_t reaches α, then the data is of the form

$$\mathbf{T} = (T_i)_{1\times N}\,, \tag{2}$$

for the N times (e.g. or RT observations) that the SW distribution describes (T is also known as the *first passage time* of the BMP). Parameter θ functionally accounts for these aspects external to the RWD by shifting all values of t by a constant, in which the starting point of the accumulation process, $X_0 = 0$, instead becomes, $X_\theta = 0$. While θ shifts the distribution from the left, note that its effect, mathematically, is equivalent in being able to account for external processes that occur on either side of the accumulation event.

As a Distribution Measurement Tool

While the SW and its parameters can directly describe the data in the context of a latent quantity accumulating to threshold, the SW can also serve as an objective distribution measurement tool, in which its parameters, γ, α, and θ, will directly quantify the density of the observed RT distribution.

[1]The RWD form in (1) is the same kind used by other models of accumulation: the LBA, race, and DDM, with elementary adjustments; these are specified in the Discussion.

The SW distribution with probability density function

$$f(X \mid \gamma, \alpha, \theta) = \frac{\alpha}{\sqrt{2\pi(X-\theta)^3}} \cdot \exp\left\{-\frac{[\alpha - \gamma(X-\theta)]^2}{2(X-\theta)}\right\}, \tag{3}$$

has expected value $\alpha/\gamma + \theta$, and variance α/γ^3, for $X > \theta$. The pdf is illustrated in the right plot of Figure 1, in which the distribution in black print has parameters $\theta = 200$, $\gamma = 0.08$, and $\alpha = 40$; then in different shades of grey, the figure also illustrates the outcome obtained when each of these parameters are individually adjusted in the direction that results in bigger data values (e.g. slower RTs). In each parameter adjustment, there is a unique distribution outcome: for example one can see parameter θ will give the position of the leading edge of the distribution, and shifts the entire distribution horizontally (to the right for slower RTs); then γ and α both serve to locate the central tendency within the shifted distribution; but γ is more informative for mass in the tail, and hence steepness of the leading curve (thicker tail for slower RTs); and α for the deviation centrally around the mode value, and hence normality around the mode (larger deviation for slower RTs).

While these individual parameter adjustments illustrated in the right plot of Figure 1, each provide for a unique distribution outcome, note that some of these distributions however share similar mean RTs, such as the dark and medium-grey distributions. Such can be the case in real data, when markedly different distributions, with near-equal means, are observed across experimental manipulations. Experiment manipulations with such contrasting distribution results, yet similar mean RTs, could likely cause a Type II error in classical analyses that mainly compare the means.

The advantage of the SW as a measurement tool is its ability to parse the distribution for these features by its three-parameter *decomposition of the central tendency*, in which as noted before, $E(X) = \alpha/\gamma + \theta$. In total, noteworthy advantages of the SW may include: (1) the whole distribution being fit across each experimental manipulation; (2) experimental manipulations being quantified along three kinds of distinct distribution outcomes; (3) observed RT data means being decomposed according to their distributional make-up; (4) observed means with similar values may be revealed rather as markedly different; and (5), the RT data being fit on its natural scale by the SW, with no need for an inherently-imperfect approximation to the normal. These benefits can be further supported by early works expounding the importance of accounting for the full RT distribution by Luce (1986); also some of the aforementioned benefits are explicitly discussed by Balota et al. (2008); Balota & Yap (2011) yet in the context of the ex-Gaussian, which is also an excellent distribution measurement tool, but does not have this direct correspondence to a latent accumulation process.

Utilizing the Shifted Wald

Whether one decides to utilize the SW as a distribution measurement tool, or as an accumulation model of the data, the approach of use is the same: to simply fit the distribution, which is to estimate its three parameters. It is the same approach since the parameters of the SW simultaneously describe both the shape of the RT distribution, and the data in the context of latent accumulation to threshold.

Application to Simulated Data

We developed a fitting method that combines techniques of deviance criterion minimization of observed-versus-predicted quantile distance, and maximum likelihood (ML) estimation, to fit the model parameters. The approach is summarized as follows. In the case of the SW, given a parameter value for its shape β, the other two parameters, θ and α, may be determined by closed-form ML estimators, developed as in Nagatsuka & Balakrishnan (2013). Parameter γ is then obtainable as $\gamma = 1/\alpha\beta$. An algorithm searches the near-entire space of β, and for each β, computes the model-predicted quantiles in the near-full range at high resolution (e.g. 100 equally-spaced quantiles between the .02-.98, or .001-.999 range depending on choice of fitting outliers in the data).[2] The parameter set that leads to the smallest absolute difference in the observed-versus-predicted quantiles is selected as the fit.

We have found the SW in the context of this method, to be robust in the recovery of parameters during cases of both small numbers of observations $N = 50$, as well as large $N = 1000$; the fitting procedure finishes on the level of seconds using standard computing technology, and can be simply performed via R or MATLAB. The following table contains the simulation recovery results, which are the average Pearson r correlations between the model fit and generating parameters across 1000 recovery simulation trials; each row corresponds to the recovery of a different data set size (e.g. number of observations), N.

Table 1: Parameter Recovery, Average Pearson Correlations

Observations	γ	α	θ
$N = 1000$	0.99	0.95	0.99
$N = 500$	0.98	0.93	0.99
$N = 250$	0.97	0.89	0.99
$N = 125$	0.94	0.82	0.98
$N = 50$	0.88	0.68	0.97
$N = 15$	0.72	0.47	0.92

In addition, the fits also matched the observed data quantiles very well. Given the desirable performance of the method, we utilize the approach to fit the real data.

Application to Real Data

In this section, the fitting approach is demonstrated on a data set involving a visual search (VS) task by baboons of mixed ages, collected by Goujon & Fagot (2013); and the results

[2]For more information, see Anders et al. (in review).

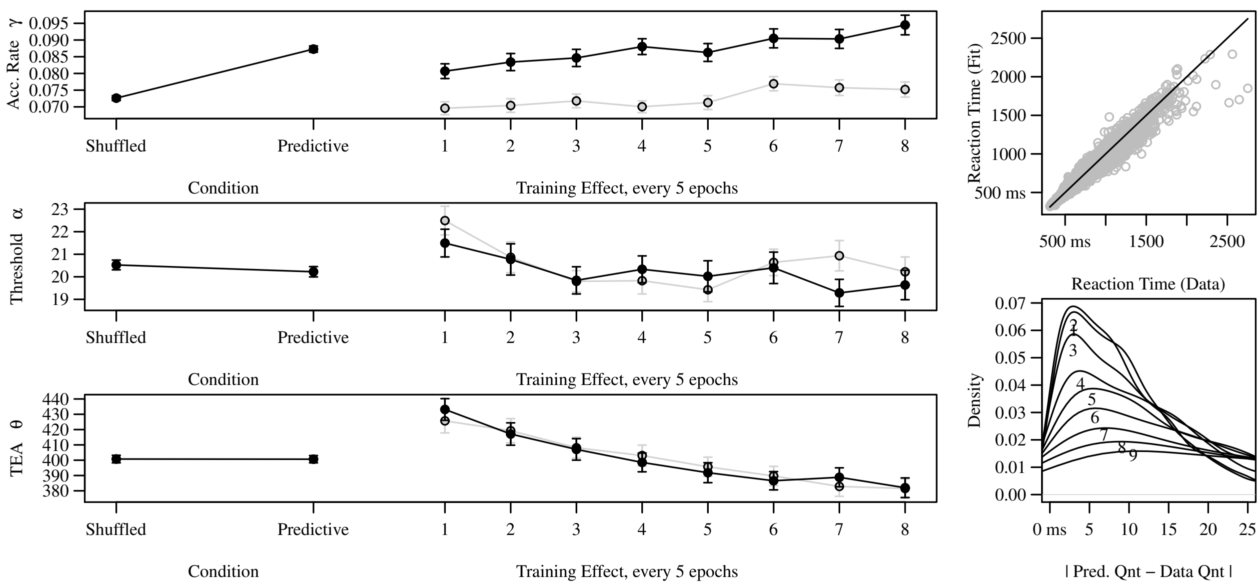

Figure 2: The SW fit to the VS task by baboons: (left) mean parameter values with bars representing standard error of the mean grouped by experiment factor, $N_C = 1080$ and $N_{Tr} = 135$, in which the grey dots show the training effect for the non-predictive (shuffled) condition, and the black for the predictive condition; (top-right) fitted-versus-observed, quantile-quantile plot for the $N = 2,158$ distributions fit for all deciles; (bottom-right) the distribution of residuals for the nine deciles across the distributions.

will be presented in the vocabulary of the SW as a latent accumulation model for the data. The experimenters explored an animal model (via baboons) of statistical learning mechanisms in humans, specifically the ability to implicitly extract and utilize statistical redundancies within the environment for goal-directed behavior. Twenty-five baboons (species *Papio papio*) were trained to perform a VS task with contextual cueing. The task consisted of visually searching for a target (the letter “T”) that was embedded within configurations of distractors (letters “L”), which were either arranged predictively to locate the target (hence a contextual cue), or non-predictively (shuffled, without a cue).

As organized by the original researchers, there are three meaningful partitions: the $C = 2$ predictive vs. non-predictive contextual cue conditions; the $E = 40$ time-points (epochs) to observe training effects, in which every unit step in E consists of 5 blocks (each block contains 12 trials, and thus each E contains 60 trials); and the $B = 27$ individual baboons. These three meaningful factors provide for $N = 2158$ separate distributions to each be individually fit by the SW. The average distribution length (number of observations) is $\bar{L} = 30$, with standard deviation, $\mathrm{SD}(L) = 1.10$.

Figure 2 provides the results of the analysis on the baboon VS task. The left column of three plots respectively contains the means, and their standard errors, of the model-fit measurements of three SW parameters: γ, α, and θ, grouped by experiment factor: condition ($N_C = 1080$) and training effect, in which the levels are averages of every five proceeding epochs, to simplify the illustration ($N_{Tr} = 135$ per each of two conditions). Beginning with consideration of condition, the model clearly isolates the effect of condition (non-predictive vs. predictive) on a single parameter, the accumulation rate of signal strength (or target detection), γ; note that the standard errors of the mean in this case are in fact too small to be seen in the plot. The other parameters, α and θ, which in this task might be respectively interpreted as a certainty criterion before responding, and mechanical response/visual processing RT speed, showed no substantive change across conditions.

Next, the analysis of training effects over time are displayed for each contextual cue condition: the predictive condition in black points, and the non-predictive (shuffled) condition—which provides little information (e.g. cues) to learn from while doing the task–in grey points. The training effect appears to adjust each of the parameters over time in a way that supports faster RTs, yet interestingly in different ways. Most notably, the TEA parameter, θ, for mechanical/perceptual RT processes, benefits equally by training across epochs during both conditions–which is a rather plausible finding—as does the response caution / signal criterion parameter, α. In contrast, there is a marked difference across conditions in the benefit rate of the signal accumulation parameter, γ, by training.

Furthermore, while each of the SW parameters appears to be modulated by training, they differ in their rate of change over time, and their onsets/magnitudes of diminishing returns. For example, γ appears to benefit in a consistently-increasing linear fashion from levels 1 to 8; while α and θ speed benefits occur in uniquely different curvi-linear fashions, with different diminishing or zero-return onsets, respectively near training points 5 and 7.

The right column of plots in Figure 2 provide model goodness-of-fit checks to verify if the observed data quantiles are appropriately fit by the SW. The top plot contains the deciles of all $N = 2158$ distributions fit with the SW; as one can see, nearly all of the fits match the observed deciles well in a corresponding $x = y$ fashion, with very few outliers. The bottom plot provides the distribution of residuals for each of the nine deciles across the 2158 cells fit; here it is shown that most of the deciles are similarly well fit, with a slightly larger variance for the deciles 7-9, which tend to also hold increasingly larger RT magnitude and variation in the observed data.

Discussion

The utility of the SW distribution, to serve as a cognitive model for certain response tasks by describing the data in the context of accumulation to threshold, as well as its usefulness as an objective measurement tool for RT distributions, was presented. Noteworthy and unique aspects of the SW, which set it apart from other distributions that may be used as RT distribution measurement tools, include its flexibility to accommodate a number of distribution shapes; its three-parameter decomposition of the mean, each parameter accounting for a distinct distribution outcome; its ability to be fit well during cases of few observations; and most distinctively its unique ability to also describe the data via accumulation to threshold.[3]

Important clarifications can be made to resolve confusions between the SW distribution, particularly its accumulation model characteristic, and more complex accumulation models such as the DDM, race and LBA models. Firstly, the SW distribution is the only model of the three in which its parameters directly quantify the distribution of RTs, and simultaneously directly describe the RT data in the context of a latent quantity accumulating to threshold. Secondly, it always consists of only one accumulator modeling the response process, with one threshold.

On the latent accumulator aspect of the SW, there are only minor modifications which will deliver the researcher to one of the three other prominent models: the DDM, race, and LBA. Each of these three models have the same kind of accumulator as in the SW: the DDM instead has two thresholds: a lower and upper, to model two characteristic outcomes; and hence allows for negative drift rates, e.g. $\gamma < 0$, to allow substantial observations on the lower boundary. The race model has multiple instances of the same accumulator as the SW, to model any number of characteristic responses, in which the first accumulator that reaches the threshold wins. The LBA has this same property of the race model, except the latent quantity accumulates in a constant linear fashion (known as a "random ray"), rather than as drift with random noise.

[3] Indeed other measurement distributions (e.g. ex-Gaussian, Gumbel) may also provide excellent utility or fits of RT data. However their principal difference from the SW, is they do not possess the ability to also describe the data by accumulation to threshold.

Thus all of these approaches are indeed very closely related. For example, the SW and DDM could be said to constitute the very same supra-model: they both stem from the same family process, the Wiener process, and as mentioned, arise from only subtle differences in parameter values (see respectively Chapter 3, and pages 8–24, Chhikara, 1988; Gerstein & Mandelbrot, 1964; Jones & Dzhafarov, 2014, for more information); in which some parameterizations of the Wiener process result in a closed-form probability density function (e.g. the SW), while others will not (e.g. the DDM). They are hence simply nested models, both using the same kind of RWD, or Brownian motion process designed in (1). Therefore in the context of an appropriate data application, an attack or critique on the elements of one of these models, such as the validity of the cognitive interpretation of this latent Brownian motion process, may be considered an attack on all three models.

A concrete issue of practicality however, worth mentioning between simple models, such as the SW, that have one accumulator and one threshold for the observed response, and more complex models that seek to have separate accumulators, and/or thresholds for every response option, is the large benefit of the SW in the context of the *limitation of the data*. More specifically, the limitation of the number of observations available in the data, per experimental manipulation and per response alternative, can be a problem that is exacerbated much more quickly as one increases in the extra numbers of accumulators and/or thresholds that more complex models have. For example in our baboon data application, the depth that we explore the experimental manipulations, estimating individual parameters for each combination of them, by participant, is a resolution that would not have been appropriate for the other more complex models. This is because there were insufficient amounts of observations for each response alternative, per experimental manipulation, to drive the estimation of the other models' extra parameters that arise from additional accumulators / thresholds; these models would be attempting to model data, with far too many missing observations. Thus it is important to take into account the number of observations per data cell sought to be analyzed: when selecting (1) an accumulator model variant, and (2) the depth that one parameterizes the model across cells; e.g. having enough observations (such as > 30) for each extra response option modeled, per cell fit.

The limitation of the SW, for being applied to data with many response alternatives observed per experimental manipulation, is it lacks the ability to serve as a complete generative model for the full data. For example, considering data with substantial amounts of both corrects and errors, the SW can be applied separately to the corrects and errors. Here it may serve as a distributional measurement tool to quantify distribution differences across conditions, and/or deliver a latent accumulation account across experimental manipulations, conditional on the respondent providing that observed (correct or error) response. However in this case, the SW cannot serve as a full generative model, for example to produce near the same number of observed number of corrects and errors, by only knowing the parameters alone, and not how many were corrects and errors were observed in the first place. In contrast, a model such as the DDM, race, or LBA, can not only serve to account for differences between the experimental manipulations, but also as a complete generative model for the data, by having the *a priori* probability of a correct or error response, already pre-coded in the model, by being in the respective drift rates for each experimental manipulation; and thus are excellent tools for these multi-response option cases.

Thus in each model having its unique assumptions, benefits, and restrictions, it is up to the researcher to select the model(s) that best suit his or her research aims within the particular application. While there are certainly appropriate situations and data that could considerably benefit from a SW analysis approach, currently there are very few publications in the psychological literature that utilize the distribution. We hope to have advocated the distribution's use, as well as to have facilitated a deeper understanding of the SW, and its position in the context of accumulation modeling.

Acknowledgments

We acknowledge funding by the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant agreement n° 263575), and the Brain and Language Research Institute (Aix-Marseille Université : A*MIDEX grant ANR-11-IDEX-0001-02 and LABEX grant ANR-11-LABX-0036). We thank the "Féderation de Recherche 3C" (Aix-Marseille Université) for institutional support, as well as Joël Fagot, for making available his data.

References

Anders, R., Alario, F.-X., & van Maanen, L. (in review). The shifted Wald distribution for response time data analysis, .

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 514.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science*, *20*, 160–166.

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 and Language*, *59*, 495–523.

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Chhikara, R. (1988). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications* volume 95. CRC Press.

Gerstein, G. L., & Mandelbrot, B. (1964). Random walk models for the spike activity of a single neuron. *Biophysical Journal*, *4*, 41–68.

Goujon, A., & Fagot, J. (2013). Learning of spatial statistics in nonhuman primates: Contextual cueing in baboons (papio papio). *Behavioural Brain Research*, *247*, 101–109.

Heathcote, A. (2004). Fitting Wald and ex-Wald distributions to response time data: An example using functions for the s-plus package. *Behavior Research Methods, Instruments, & Computers*, *36*, 678–694.

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

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*, 1.

LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, *27*, 375–396.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. 8. Oxford University Press.

Nagatsuka, H., & Balakrishnan, N. (2013). A consistent method of estimation for the parameters of the three-parameter inverse Gaussian distribution. *Journal of Statistical Computation and Simulation*, *83*, 1915–1931.

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

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

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

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

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1280.

Van Maanen, L., Grasman, R. P., Forstmann, B. U., & Wagenmakers, E.-J. (2012). Piéron's law and optimal behavior in perceptual decision-making. *Frontiers in Neuroscience*, *5*.

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

Van Zandt, T. (2002). Analysis of response time distributions. *Stevens' Handbook of Experimental Psychology*, .

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*, 830.