

Modeling Response Time with Power Law Distributions

Zhiyuan Liu, *Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221-0011*

John G. Holden, *Department of Psychology, CAP Center for Cognition, Action, and Perception, University of Cincinnati, Cincinnati, Ohio 45221-0376*

M. Dashti Moghaddam, *Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221-0011*

R. A. Serota, *Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221-0011*

RUNNING HEAD: Modeling Power-Law Distributions

Correspondence should be addressed to: John G. Holden, Department of Psychology, University of Cincinnati, Cincinnati, Ohio 45221-0376, E-mail: john.holden@uc.edu, Phone: 513-556-5519.

Preprint of article to appear in: *Nonlinear Dynamics, Psychology, and Life Sciences* ©
This is not the document of record, please cite the published article.

Abstract

This article overviews several contemporary models that assume power law scaling is a plausible description of the skewed right tails that are typical of response time distributions. The properties and markers of these distribution functions have implications for cognitive and neurophysiological dynamics. The power law hypothesis suggests studies should collect larger samples, and that analyses may combine individual subjects' data into a single set for a distribution-function contrasts. Techniques for contrasting response time measurements are illustrated on data from a previously published study comparing the performance of children diagnosed with dyslexia and a group of age-matched controls in flanker, color naming, word naming, and arithmetic performance.

KEYWORDS: cognition, dyslexia, response time, scaling, power law

INTRODUCTION

Response time (RT) is the elapsed interval between the presentation of a stimulus and the execution of a response in a laboratory-based cognitive task. Response times are widely used in basic cognitive science and routinely used in neuropsychiatric assessments for reading disabilities and certain cognitive impairments (e.g., Ancelin, Artero, Portet, Dupuy, Touchon, & Ritchie, 2006; Epstein, Brinkman, Froehlich, Langberg, Narad, Antonini, & Altaye, 2011; Osmon, Kazakov, & Kassel, 2018; Tse, Balota, Yap, Duchek, & McCabe, 2010). To date, there is little scientific consensus regarding the fundamental statistical properties of response time distributions themselves (e.g., Holden, Van Orden, & Turvey, 2009; Jones & Dzhafarov, 2014; Luce, 1986). In light of their widespread use, an accurate statistical portrait of response time distributions would be extremely useful. We introduce several descriptions motivated by dynamic systems and complex network theory.

This article is focused on response time probability density functions (PDFs) that entail strongly skewed power-law tails. Several groups have proposed that the characteristic positive skew in the right tails of response time distributions is consistent with a power-law decay function (e.g., Holden, & Rajaraman, 2012; Holden et al., 2009; Ihlen, 2013; Moscoso Del Prado Martin, 2008; Sigman, Etchemendy, Fernandez Slezak, & Cecchi, 2010).

After first introducing techniques for modeling power-law distributions, we apply a selection of these models to a published data set that compares the perceptual performances of a group of children diagnosed with dyslexia and their age-matched controls (Holden, Greijn, van Rooij, Wijnants, & Bosman, 2014). The original article applied the fractal concept of distribution scaling to the performances. The goal was to evaluate a hypothesis that two groups differ in degree, rather than in kind. The present article revisits that basic question with more sensitive statistical techniques.

We tentatively assume that response time PDFs are at least approximately stationary – they stabilize near a steady state for a given cognitive activity. This assumption is tentative because of reports that time-series of response times express $1/f^\alpha$ noise – long-range correlations and fractional dimensionality (Gilden, 2001; Van Orden, Holden, & Turvey, 2003; 2005). This pattern is also called pink noise, fractal time, and $1/f^\alpha$ scaling — the term we adopt going forward. $1/f^\alpha$ scaling unfolds over time and offers a window into the significance of the power-law behavior in the tails of response time density functions, as we now explain.

Lessons from Temporal Scaling

Reports of monofractal $1/f^\alpha$ (e.g., Wijnants, 2014) and even multifractal scaling (e.g., Cavanaugh, Kelty-Stephen, & Stergiou, 2017) in sequential signals derived from human activity and physiology are now common. The essential characteristic of these scaling relations is a self-similar pattern of fluctuation in the measured signals: fluctuations observed in runs of several successive measurements are statistically echoed in progressively larger and larger runs of observations (see Gilden, 2001; Holden, 2005; Riley & Holden, 2012 for overviews). These nested, scale-free patterns of fluctuation are associated with complex systems that self-organize their temporal and spatial behavior (Jensen, 1998).

In human activity, the basis of $1/f^\alpha$ scaling is becoming clear. Consider a conventional cognitive task that presents successive trials requiring participants to respond with an elementary utterance or button-press indicating an identity or judgment about each presented stimulus. The duration of each response time gauges the time required for mind and body to register and organize a coherent response. However, the temporal scaling revealed in response time trial-

series unfolds on slower time-scales than the trial-by-trial pace of the study, and the cognitive acts that are its focus (Van Orden et al., 2003).

Instead, the $1/f^\alpha$ scaling captured in response time tasks emphasizes fluctuations in the physiological and bodily processes that *support* ongoing cognitive activity. Illustrative examples include interrelated metabolic fluctuations in respiration, heart rate, and consequent blood and brain oxygenation levels (Gonzales, Hessler, & Amazeen, 2010). Blood glucose levels are similarly intertwined with insulin levels and fluctuate on slightly slower time scales. Likewise, intrinsic $1/f^\alpha$ scaling is expressed in neurophysiological signals over time scales ranging from milliseconds to at least 20 minutes (e.g., EEG and MEG, Linkenkaer-Hansen, Nikouline, Pavla, & Limoniemi, 2001; PET, Bullmore et al., 2001; He, 2011; fMRI, Fox, & Raichle, 2007). Animal and human studies demonstrate that gross levels of major neurotransmitters such as GABA, glutamate, dopamine, their metabolites, and their receptor binding affinities all fluctuate on time scales ranging from just seconds to hours, days, and beyond (Casteñeda, Prado, Prieto, & Mora, 2004; Wirz-Justice, 1987). The intertwined states of these and countless other biological and physiological variables have the potential to contribute to performance variability over time.

Biological and physiological fluctuations reflect ongoing multiscale cooperative and competitive interdependencies required to balance extrinsic environmental demands and intrinsic physiological requirements. Apparently, $1/f^\alpha$ scaling reflects this metastable coordinative activity (Kello, Anderson, Holden, & Van Orden, 2008; Kello, Beltz, Holden & Van Orden, 2007). Physiological variables rarely evolve randomly or instantaneously, but any number of them can subtly impact an individual's ongoing performance. As such, sequences of discrete, elementary cognitive acts carry the history of these influences on their backs, so to speak (Van Orden, Hollis, & Wallot, 2012). Rather than homeostasis, Bassingthwaighe, Liebovitch, and West (1994) adopt the term homeodynamics to emphasize the constantly evolving multiscale nature of physiological variables.

All this coordinative activity is both accommodated and constrained by layers and layers of entrainments, interdependences, and rate-limiting interactions (Buzsáki, 2006; Izhikevitch, 2007; Van Orden et al., 2012). Many essential physiological, perceptual, and movement processes rely on principles of self-organization and oscillatory entrainment to become enfolded into a unified coordinative organization, or *synergy* appropriate for supporting a given human activity – known as *interaction dominant dynamics* (Amazeen et al., 1995; Amon, Pavlov, & Holden, 2018; Holden & Rajaraman, 2012; Kelso, 1997; 2012; Kugler & Turvey, 1987; Van Orden et al., 2003). The essential lesson of $1/f^\alpha$ scaling in human performance is that homeodynamic coherence in body and mind is achieved through vast networks of neurochemical, metabolic, and physiological dynamics (Bassingthwaighe et al., 1994).

Implications for Distribution Scaling

In contrast to the slower time scales emphasized by $1/f^\alpha$ scaling analyses, this article illustrates techniques for using response time distributions to characterize coordinative activity and scaling behavior closer to the time-scale of responding. It emphasizes the interval between stimulus and response in a cognitive task. In fact, conventional probability density modeling and fitting techniques allow for this perspective, as we now explain.

The time or trial-series approach to response time analysis collapses the aforementioned continuous multidimensional trajectories of homeodynamic fluctuations into a two-dimensional time-by-observation *trajectory*. Subsequently ignoring the sequential order, the time-axis, of the measurements further collapses the two-dimensional trajectory into set of one-dimensional points

that can be depicted with a conventional PDF (van Rooij, Nash, Rajaraman, & Holden, 2013). That is, the x-axis is response time, and the y-axis is the probability of observing a response time in any given time interval. Thus, conventional probability density functions portray response times as recurrent events, in a manner akin to the way Poincaré sections depict periodicity in a reduced state-space of a higher dimensional dynamic system. Of course, sampling intermittency and dimensional compression reduces dynamical resolution. However, the PDF characterization reveals the return times of a specific dynamical state that is of great interest to cognitive science: the registration of a coherent response following a relatively discrete cognitive act.

Conventional PDFs assume a range of shapes. Several psychologically relevant examples fall along a continuum from the compact and symmetric normal distribution, to the somewhat skewed or heavy-tailed exponential and lognormal distributions, and the fat-tailed inverse power-law distribution (West & Deering, 1995).

Variability in a normal or exponential distribution arises from unsystematic, additive, and memoryless sources. By contrast, distribution scaling is symptomatic of multiscale coordinative activity, spanning a wide range of time scales (van Rooij et al., 2013; Holden et al., 2009; Holden & Rajaraman, 2012). As with $1/f^\alpha$ scaling, a positively skewed PDF with a tail that decays as a power law is often symptomatic of scale-free coordinative activity. A power law relates the very small and very large measurements, since they all obey the same scaling relation. In fact, the exponent describing a distribution's power-law tail indexes the relative coherence of the coordinative dynamics supporting the measured variable, and it can be understood as a fractal dimension (Seuront, 2010; van Rooij et al., 2013).

Essentially, the scaling expressed in repeated measurements of the same cognitive act serves as a gauge of the coordinative activity supporting the act. There are a few important caveats to this statement: First, this index is most reliable in the context of the same individual performing the same activity, under the same experimental circumstances. This article seeks to generalize the inference to between-subjects research designs. Thus, we assume that different individuals accomplish identical cognitive acts using similar coordinative dynamics. One might reasonably question this assumption. Nevertheless, it is a historically pervasive statistical and theoretical assumption in psychology and cognitive science. Second, scaling exponents do not *uniquely* characterize any given human activity. Different tasks, emphasizing different cognitive acts may yield identical scaling exponents. Owing to the flexibility of human activity, only comparisons of the relative magnitudes scaling exponents derived from carefully matched groups, or well controlled experimental contexts make comparative sense.

In what follows, our working hypothesis is that performances yielding RT densities with particular parameters, observed in the context of a particular cognitive activity, require the support of a network of particular perceptual, cognitive and neurophysiological processes (Holden et al., 2012; Holden, Ma, & Serota, 2013; Holden et al., 2014). At any given point in time, fluctuations may arise in the organization of this network and its parametric quantities. This is true both for individuals performing a given task, and in contrasts across separate groups of individuals in different experimental conditions.

Assuming comparable process networks that express fat-tailed behavioral measures implies individual participants' performances can be viewed as samples or realizations of random variables that are representative of the potential network states supporting a given cognitive activity. In this situation, combining observations across participants may clarify the potential range of parametric variation that one must expect from the task, condition, or distinct participant groups, and thus improve the representativeness of statistical analyses that are derived

from the distribution of observations.

DATA AGRIGATION SIMULATION

If a power-law model provides a plausible description of an empirical distribution, then several statistical and methodological implications follow. Traditionally, experimental sample sizes are derived from either formal (e.g., power analysis) or informal Gaussian rules-of-thumb (e.g., 30 or more data points yields reasonable mean (location) estimates). Robust parameter estimates for power-law models, and especially the scaling exponents describing their skewed tails, require much larger samples. Sample sizes can easily be ramped up in some populations and experimental paradigms, but other populations and paradigms entail inherent limitations. For example, studies relating to populations that involve simultaneous imaging or EEG measurements, health impairments, children, or the aged, often face significant time-on-task limitations. This simulation tests the viability of a technique to compensate for smaller samples by combining data across participants in a given experimental cell. The simulation discussion section offers additional recommendations for experimental designs that quantify distribution scaling, and for interpreting different patterns of change in power-law model parameters. Once these methodological techniques are in place, we illustrate their use on a data set that contrasts the performance of a group of children with dyslexia and a group of age-matched normally reading children.

Overview of Potential Power-Law Models

Previously, we examined the skewed tails of the probability density functions of response times (Holden & Rajaraman, 2012; Holden et al., 2009; Ma, Holden, & Serota, 2015). Power-law scaling prevailed as a more plausible description of empirical RT distribution's tail behavior than models adopting exponentially decaying tail behavior or even exclusively lognormal behavior.

In in this section, we first introduce three candidate distributions that implement power-law tails: The Lévy alpha-stable distribution (Stable), the generalized inverse gamma distribution and the “Cocktail” lognormal-Pareto mixture distribution. Following our introduction to the power-law distribution models, we use both the generalized inverse gamma and the lognormal-Pareto distributions to introduce fat-tailed research design practices and other methods for aggregating response time observations to support statistical inference. Finally, we illustrate the use of these selected techniques on a published data set that contrasted a sample of children diagnosed with dyslexia and matched age-appropriate readers.

To begin, we summarize the parametric properties of three potential model distributions with power-law tails. Notably, model parameters that determine the origin or location of a distribution on the x-axis are referred to as *location* parameters. *Scale* parameters control the relative spread of the distribution. Higher-order asymmetries, such as skew and kurtosis, are controlled by *shape* parameters. More detailed explanations of the role of the location, scale, and shape parameters appear in the context of the definition of distribution rescaling in the simulation discussion section.

The Stable Distribution. With the exception of $\alpha = 2$, when the stable distribution reduces to a normal distribution, the stable distribution (Nolan, 2015), has a power-law tail, such that

$$S(x; \alpha, \beta, \mu, \sigma) \sim |x|^{-(1+\alpha)} \quad (1)$$

Here stability is bounded by a shape parameter: $0 < \alpha < 2$, defined below. Skewness is bounded as $-1 \leq \beta \leq 1$ and is the second shape parameter of the distribution. The parameters σ and μ control its scale and location, respectively. In general, the stable distribution does not

have a closed-form expression, with the notable exceptions of the Cauchy and Lévy distributions. Instead, the stable distribution is described by a closed-form characteristic function and via parametric forms.

The two important properties of the stable distribution are as follows: First, the sum of two independent and identically distributed (i.i.d.) variables distributed as $S(x; *, *, *, \sigma)$, is distributed with $S(x; *, *, *, 2^{1/\alpha} \sigma)$. That is, the linear combination of i.i.d. variables result in the same distribution, up to a rescaled scale parameter – this is the definition of stability. Second, the sum of i.i.d. variables, whose PDF has an $|x|^{-(1+\alpha)}$ tail, $0 < \alpha < 2$, tends to $S(x; \alpha, 0, 0, \sigma)$ as the number of variables in the sum grows, by the generalized central limit theorem.

Ihlen (2013) proposed the stable distribution for fitting of RT distributions. His goal was to preserve the conventional assumption that cognitive performance originates in the output of sums of distinct cognitive processes. Ihlen reasoned that if the outputs of individual cognitive processes conform to the stable distribution, and are combined as sums in the resulting responses, then RTs themselves should be distributed as a stable distribution. Ihlen demonstrated that the stable was a better fitting distribution than the ex-Gaussian or shifted Wald for a large catalogue of RT data. However, no other fat-tailed distributions were evaluated, and Ihlen stated that other fat-tailed distributions might yet provide better fits.

While the stable distribution is an attractive model due to its properties above – namely, power-law tails and additivity – its two main drawbacks are that its scaling exponents, $(1 + \alpha)$, are limited to the range between 1 and 3, and the only permitted value of β is $\beta = 1$, otherwise the variable of the distribution is undefined for positive values, as is required by physically and biologically meaningful RT models dealing with elapsed time intervals greater than zero. Both the recinormal and Fieller's distributions, whose scaling exponent is limited to a single value of 2 suffer from similar limitations (see Moscoso Del Prado Martin, 2008). Essentially, the permitted values of all three of these model exponents are contradicted by the observed exponent values in several large-sample RT studies (see, for instance, Holden et al., 2009; 2012; Ma et al., 2016). As such, we do not implement them as part of our illustrative modeling.

Generalized Inverse Gamma Distribution. The application of generalized inverse gamma to RT modeling is motivated by its connection to the generalized Bouchaud-Mézard network model (Bouchaud, & Mézard, 2000) – and the network's potential implications for brain dynamics. The Bouchaud-Mézard network is controlled by stochastic differential equations, that entail a stochastic activation-inhibition or “birth-death” dynamic (Ma et al., 2016, Ma, Holden, & Serota, 2013; Ma, & Serota, 2014). The generalized Bouchaud-Mézard network model is given by Eq. 3 in Ma, Holden, and Serota (2013) and describes exchanges between nodes in a partially connected complex network, subject to stochastic perturbations, which are proportional to the current node values. The network entails a connection strength parameter and connection numerosity parameter. Together, both parameters control the scaling exponent of the network's output distribution.

We do not propose direct architectural or functional links between the Bouchaud-Mézard network and the neurophysiological networks involved in human performance. However, the transitivity of network dynamics suggests the viability of using network simulations to test basic hypotheses regarding the potential influence of network connectivity parameters on performance. This objective is in-line with our hypothesis of a close correspondence between the shapes of observed RT distributions and the network dynamics of supporting cognitive and neurophysiological dynamics. The three important parameters in a generic Bouchaud-Mézard

network model are 1.) the exchange strength J between the nodes, 2.) the magnitude of stochasticity σ , and 3.) the phenomenological parameter γ describing the degree of connectivity in the network: $\gamma=1$ is a fully connected network and $\gamma \rightarrow 0$ is a vanishing number of connections. These parameters may be relevant to modeling or simulating connectivity and interactions among brain regions.

The generalized inverse gamma distribution μ parameter defines its onset, so the generalized inverse gamma PDF is zero if $x < \mu$, while for $x > \mu$ it is given by:

$$GIGa(x; \alpha, \beta, \gamma, \mu) = \frac{\exp\left(-\left(\frac{\beta}{x-\mu}\right)^\gamma\right) \gamma \left(\frac{\beta}{x-\mu}\right)^{\alpha\gamma+1}}{\beta \Gamma(\alpha)} \quad (2)$$

In Eq. (2), α and γ are shape parameters, β is a scale parameter, μ is the location parameter, and Γ indicates the Gamma function. The generalized inverse gamma shape parameters relate directly to the Bouchaud-Mézard model parameters as $\alpha\gamma = 1 + J/\sigma^2$ and $\beta\theta(\gamma)\gamma^{-1} = J/\sigma^2$, where $\theta(\gamma)$ varies monotonically between $\theta(0) = 1 + J/\sigma^2$ and $\theta(1) = 1$. Except for μ , all generalized inverse gamma parameters must be greater than zero and, in the context of RT fitting, μ must also be a positive number. For large x , where $x - \mu \gg \beta$, the generalized inverse gamma distribution exhibits a power-law tail given by

$$GIGa(x; \alpha, \beta, \gamma, \mu) \sim x^{-(1+\alpha\gamma)} \quad (3)$$

Unlike the stable distribution, there is no a priori upper bound on the tail exponent of generalized inverse gamma.

Lognormal-Pareto Mixture Distribution. In the cocktail model (Holden et al., 2012; Holden et al., 2009), the lognormal-Pareto mixture distribution is given by Eq. 4. While using slightly different notations, this form is equivalent to that in Holden et al. (2012):

$$LNP(x; \mu, \sigma, \alpha, \rho_p) = \begin{cases} \frac{\rho_<}{C\sqrt{2\pi\alpha}} e^{-\frac{|\ln x - \mu|^2}{2\sigma^2}}, & x \leq x_p \\ \frac{\rho_>}{(1-C)\sqrt{2\pi\alpha}} e^{-\frac{|\ln x - \mu|^2}{2\sigma^2}} + \frac{\rho_p \alpha x^\alpha}{x^{\alpha+1}}, & x \geq x_p \end{cases} \quad (4)$$

Here $C = \frac{1}{2} \left(1 + \text{Erf} \left(\frac{\alpha\sigma}{\sqrt{2}} \right) \right)$, where *Erf* is the error function, α , σ and ρ_p are the shape parameters and x_p is the scale parameter. The other three parameters, $\rho_<$, $\rho_>$ and μ are determined from the condition of normalizability of lognormal-Pareto ($\rho_< + \rho_> + \rho_p = 1$) and continuity of lognormal-Pareto and its derivative at x_p ; in particular, μ and x_p are related by $x_p = \exp(\mu + \alpha\sigma^2)$.

The choice of the LN front end is motivated by multiplicativity. For large x ,

$$LNP(x; \mu, \sigma, \alpha, \rho_p) \sim x^{-(1+\alpha)} \quad (5)$$

As is for generalized inverse gamma, lognormal-Pareto does not have an upper bound for

the tail exponent. However, unlike the stable distribution and the generalized inverse gamma, which are infinitely differentiable at all points, lognormal-Pareto's derivatives above first are discontinuous at x_p .

The lognormal-Pareto is motivated by the characteristic dynamics of coordinative synergies, such as those indicated by $1/f^\alpha$ scaling in human activity, called interaction dominant dynamics. As previously noted, individual RT measurements return discrete samples of the temporal evolution of continuously evolving dynamic trajectories. Thus, an RT distribution yields a collection of samples originating from a broad range of dynamic trajectories. Some samples reflect more organized and coherent dynamical states, others reflect less organized, less coherent dynamical states. More skilled performances tend to originate in more coherent coordinative dynamics. Less skilled performances likely rely on less coherent dynamics. Since recurrent dynamical systems operate on their own outputs, more constrained dynamics will yield lognormal variability, resulting from proportional interactions among supporting processes. Past a certain point, the relative instability of less coherent dynamics amplifies the role of recurrent feedback and noise, broadening the distribution, and yields power-law behavior. Thus, lognormal and power-law behavior anchor two ends of a continuum that are bisected by a qualitative (phase) transition from the more stable lognormal behavior to more variable inverse power-law behavior (see Holden, & Rajaraman, 2012; Holden et al., 2009; van Rooij et al., 2013).

The transition from lognormal to power-law behavior is probabilistic, and occurs within the time course of individual trials. On any given trial, available constraints arising from task and stimulus, the individual's relevant skill set, and the states of the physiological variables supporting the performance entail a potential to influence the relative coherence of performance dynamics. The result, over the course of a typical laboratory study, is a mixture of lognormal and inverse power-law samples (Holden et al., 2009; 2012).

Simulation Method

Simulation procedures. To contrast the behavior of individual and aggregated parameter estimates we conducted the following numerical simulation using the generalized inverse gamma model as an illustrative candidate distribution:

(1) We fit the by-group aggregated RT data from each of the two groups with the generalized inverse gamma to obtain the initial parameter estimates of α_0^i , β_0^i , γ_0^i , μ_0^i , where $i = C, D$, indicating either the Control and Dyslexic group respectively. The control group was initially comprised of 560 trials from each of 20 participants, the dyslexia group was comprised of 560 trials from each of 23 participants. However, trials resulting in errors were eliminated for both groups, and to equate the simulation samples, we simulated the C and D groups using 20 synthetic subjects, each of which was comprised 560 simulated observations, for each of the two groups. The values in Table 1 are the result of those fits.

(2) We used the parameter estimates of α_0^i , β_0^i , γ_0^i , μ_0^i , to generate group-size random variates from the model distribution (i.e. 560 obs. X 20 subs. = 11,200 samples). These synthetic group variables served as simulated counterparts for the empirical RT data used to generate parameters for the control (C) and children with dyslexia (D) groups in step (1).

(3) Next, we conducted a bootstrap procedure (Efron, & Tibshirani, 1994) to estimate standard errors and confidence intervals for the parameters of the simulated variables α^i , β^i , γ^i , μ^i for the C and D groups. Of course, there are other ways to generate parameter variability – we chose the bootstrap to parallel the by-group RT fitting procedure described in step (1). On each of 1000 bootstrap replications, both synthetic group-sized C and D distributions were randomly

resampled with replacement, and the parameters resulting from the fit were retained in the bootstrap distribution.

(4) Next, we generated synthetic individual-subject sized random variables comprised of 560 samples, where the parameters of these synthetic distributions were randomly sampled from a normal distribution of potential generalized inverse gamma parameter values whose mean was the real data parameter from step (1) and SD obtained in step (3). This mimicked the variability at the level of individual participants, as 560 synthetic data points were generated from a random sample of α , β , γ , and μ parameters, twenty times, representing the participants comprising both conditions. These subject sample-size distributions were then fit individually to obtain the mean, SD, and range of the resulting parameters for comparison with the group sample-size distributions.

(5) We then compared the parameters from the group-level and subject-level variables with the actual RT data parameters for the grouped and individual participants.

The simulation outcomes illustrate this approach for the Holden et al. (2014) Arithmetic study, which contrasted simple addition performance children for with and without dyslexia. Participants pressed one button to indicate that individually displayed simple addition sums, together with an answer, were correct (e.g., $3 + 6 = 9$) and another button to indicate incorrect sums (e.g., $4 + 3 = 2$). The answers to all the sums were always below 10. The Arithmetic data was chosen as the simulation test-bed due to the heterogeneity entailed in its relatively broad RT densities, relative to the remaining tasks.

Using the generalized inverse gamma as a candidate distribution, 20 simulated subject-level variables of 560 samples each were generated, as in steps (1)-(4) above. To illustrate the results, we use two measures of variability – the tail scaling exponent, computed from the generalized inverse gamma parameters as $(\alpha\gamma + 1)$ in Eq. (3) and the distribution half-width (HW), defined as the width of the distribution along the line drawn at the half-height of the PDF's maximum, which is the modal value of the PDF (MPDF). Both variability measures depend on the distribution's shape parameters.

We adopted the HW as a variability measure, rather than SD, because a PDF's power-law tail strongly biases the latter. Both the HW and power-law scaling exponent depend on at least one shape parameter. Unlike the simple dependence of the latter (see Eqs. 3 and 5), the parametric dependence of HW can be quite complex. One can think of the HW as a width of the distributions, which is ordinarily characterized by root-mean-square of the distribution. HW is proportional to the scale parameter and has a complicated dependence on shape parameters that define exponents of power-law tails, $\alpha\gamma + 1$ and $\alpha + 1$ for the generalized inverse gamma and lognormal-Pareto models respectively. We use HW for characterization of the width of both models. Figure 1 depicts the relationship between a PDF's modal value and its half width.

Insert Figure 1 about here

Simulation Results

The simulation goal was to determine if aggregating observations across participants' results in group distributions that are representative of those of the individuals comprising the group. From this perspective, there were two notable outcomes. First, the parameter estimates of the individual and group variables closely agreed with those of the actual RT data sets, at both the group and individual levels. Second, as one might expect, the most salient difference between both the actual and simulated group and individual parameter estimates is that grouped variables

produce much narrower distributions, and thus confidence intervals. The numerical results are summarized in Table 1 and Figs. 2 and 3. Comparisons of the CIs for the synthetic group variables with actual group data, and with the ranges for subject level variables, suggest that fitting of individual subjects, and subsequently averaging those parameters, may be less accurate, and yield statistically less powerful contrasts.

Insert Figure 2 about here

Each of the four plots in Figure 2 depict two histograms. The histograms in the Figure's upper right (A) and left (B) plots represent the half-width values for the control and dyslexia groups. In each plot, the darker histogram depicts the range of values when the half-width is computed separately for each individual participant in the group. The lighter histogram depicts the result when all the individual's data is aggregated together, and then repeatedly resampled using a bootstrap resampling procedure. In both cases, the bootstrap distribution produced more consistent and less variable results. The bottom two plots (C & D) depict the results of repeating the fitting procedure using simulated data sets, that mimic the sample sizes in terms of number of trials and number of participants, but that were derived from parametric fits of the generalized inverse gamma model. The behavior of the simulated data is qualitatively similar to the real data, which suggests the entailed statistical procedures provide a reasonable depiction of the empirical patterns.

Insert Figure 3 about here

Each of the four plots in Figure 3 depict two histograms. The histograms in the Figure's upper right and left plots (A & B) represent the scaling exponent describing the skew in each model's tail for the control and dyslexia groups. In each plot, the darker histogram depicts the range of values when the scaling exponent is computed separately for each individual participant in the group. The lighter histogram depicts the result when all the individual's data is aggregated together, and then repeatedly resampled using a bootstrap resampling procedure. In both cases, the bootstrap distribution produced more consistent and less variable results. The bottom two plots (C & D) depict the results of repeating the fitting procedure using simulated data sets, that mimic the sample sizes in terms of number of trials and number of participants, but that were derived from parametric fits of the generalized inverse gamma and lognormal-Pareto model. The behavior of the simulated data is qualitatively similar to the real data, which suggests the entailed statistical procedures provide a reasonable depiction of the empirical patterns.

Insert Table 1 about here

Table 1 compares empirical and simulated parameter estimates of the generalized inverse gamma tail exponent and half-width bootstrap distributions. Means are presented, with confidence intervals flanking each mean. Both the simulated and empirical data sets differed in a comparable manner. Aggregating the individual's data into an omnibus group variable for a given task yield less variable parameter estimates, with means that were similar to those of the simulated and actual individual's data sets. This indicates that aggregating power-law data with reasonably comparable parameter values does not impose qualitative changes in the distributions' parametric properties.

Discussion

The simulation results suggest that parameters derived from fat-tailed data distributions that are first aggregated *across* participants are less variable than those derived from averaging parameters taken from separate individual fits of each participant's data. Given the potential noisiness of performance measurements from special populations, such as children, and that each group, control and dyslexia, contributed only about 10^4 combined data points, the quantitative agreement seems to be quite good as well.

The outcome implies a potential benefit to the aggregation of individual's observations into a group-level variable. The approach could be particularly useful when specific observations exert undue influence on individual-level parameter estimates, as when sample sizes are small. Comparing the dispersion of the synthetic individual fits to that of the all-in-one fits recommends an all-in-one approach.

As we mentioned, there are alternative methods of generating the parameter variability, described in step (3). For instance, one could resample the raw RT distributions or use the Jackknife procedure (Efron et al., 1994). In this regard, it is notable that the bootstrap distributions of parameters from the simulated variables are narrower than those derived from empirical RT data, as evidenced in Figs. 1 and 2. Minimizing variability in parameter estimates is especially important for power-law distributions since larger samples are generally required than for narrower distributions such as the Gaussian or Exponential. Likewise, notable parameter variation can be expected in fits of individual participant's power-law distributions, especially for small sample sizes.

Next, we discuss several interpretable differences that can appear in the parameters of fat-tailed distributions arising from systems that are governed by the interaction-dominant dynamics implicated by power-law distributions (Holden et al., 2009; van Rooij et al., 2013). Following that, we illustrate the technique on data from the published Holden et al. (2014) dyslexia study.

Distribution Scaling Statistics and Methods

Systems expressing power-law behavior must be studied using methods that accommodate more time-dependence and inherent variability than is typically assumed by linear statistics. The theorems that dictate classical statistical practices are often stretched to their limits in fat-tailed applications. For example, $1/f^\alpha$ scaling undermines the classical ergodicity assumption, and much larger sample sizes are required for inference in the context of systems expressing power-law behavior than those expressing Gaussian behavior among uncorrelated samples (Amon & Holden, 2019). Extreme observations are so rare in systems with Gaussian output distributions that few experiments are designed to discover them, and unusual observations are reasonably treated as outliers. However, extreme observations are much more likely in systems expressing power-law behavior – but they may still be rare. Thus, it is important to incorporate study designs that accommodate much larger samples than is typical under the assumptions of linear statistical analysis.

Whenever possible, within-subject designs are recommended. If two categories of behavior are to be contrasted, pairwise yoking of stimuli often yields a design that is very sensitive for simple distribution contrasts. Each participant is exposed to yoked target stimuli that are matched on important control variables, but that differ in terms of the variable of interest. Both contrast distributions can then be comprised of measurements originating from the same individuals, on carefully paired target items that differ only in terms of the variable of interest (e.g., Holden, 2002).

However, many experimental manipulations require between-participant designs. In

decision studies, for instance, one can use an ideal strategy manipulation (Stone, & Van Orden, 1993). This design presents identical positive items but varies the categories of distractor or catch trials. Apparently, the method originated in lexical decision studies of word recognition, but it could be adapted to other decision paradigms. The basic idea behind the manipulation is to use different classes of distractor items that impact the relative difficulty of the required decision or discrimination (e.g., the signal-to-noise ratio). This design yields contrast distributions that are comprised of responses to identical target items from different individuals, who completed conditions with different distractor items.

If enough targets can be presented, both the within- and between-subject designs allow one to complete statistical contrasts at the level of both individual participants and analyses that aggregate across participants. In general, aggregated analyses are more sensitive, and likely more representative of systems that express rare events. Usually one looks for similar outcomes in aggregated analyses and individual level analyses. Indeed, Bayesian hierarchical modeling may be useful for this kind of evaluation, and we hope to pursue it a future work.

Similarly, some studies are designed to contrast separate groups of individuals, such as clinical classification studies, where a group of interest is compared to a control group. In these studies, health impairments may hamper efforts to present enough trials to individuals to yield stable individual level analyses. Given this circumstance, aggregated analyses become more important. Rather than fitting each subject individually, and then averaging the fits and other parameters over the group, it is often more efficient to aggregate observations from each participant into an omnibus distribution that is fit parametrically (e.g, Holden et al., 2018).

That said, there is a legitimate historical concern in the RT literature that individual and aggregate analyses have a potential to yield qualitatively different outcomes (see Estes & Maddox, 2005). This concern centers around the impact of averaging practices: Is it better to fit models individually, by participant, and average the resulting parameters or is it sufficient to average the participant's data and fit the average? Generally, the former is preferred. Notably, parameters that either stand alone or serve as the coefficient of a product that depends exclusively on the independent variable are unaffected by averaging (Estes, 1956). If two power law distributions are added or multiplied, the new distribution is a power law and the scaling exponent is simply the minimum of the aggregated distributions (Farmer, & Geanakoplos, 2008). The crucial point is that simple arithmetic combinations of power-law distributions yield qualitatively identical distributions, in which only the value of the exponent is impacted. Thus aggregating power-law distributions induces quantitative, but not qualitative changes in the distribution's tail behavior.

Our analyses of the dyslexia data, below, correspond to the simulation and combine individual participant's raw scores into omnibus distributions that are subsequently used for parameter estimation. This practice does not impose arithmetic operations, and assumes the individual datasets entail parameters that are variable, but representative of a power-law model. We followed the Estes and Maddox (2005) recommendation to complete side-by-side simulation and real data analyses to test the relationship and validity of individual and omnibus fitting for the power-law models discussed earlier. In general, we found the relationship between the individual and omnibus empirical distribution parameters is very similar to those expressed by the simulated individual and omnibus distributions. This result illustrates the utility of using omnibus parameter estimation for power-law models when circumstances demand it.

Parametric rescaling. Many natural systems are iterative in that they operate on their own outputs. For instance, new tree branches originate in existing branches. New river erosion

originates in an existing riverbed. Similarly, many dynamical systems operate on their own outputs, such as iterative “neural” networks often used to model cognitive performance. Iterative dynamics commonly produce proportionally rescaled outputs. Thus, self-similar rescaling is symptomatic of many real and synthetic fractal systems.

Rescaling hypotheses are derived directly from the behavior of iterative systems. If one set of system outputs is a rescaled copy of another, presumably the same essential system dynamic is in place, but its activity is dilated or contracted either in time or in space. By extension, rescaling in related cognitive performances suggests a similar dynamic organization is in play, but more detailed perceptual or cognitive relationships had to be resolved before a response could be organized (also see Holden, 2002; 2013; Holden et al., 2014; van Rooij et al., 2013).

Suppose one uses an identical experimental method to collect RT data from two distinct groups of participants – for instance, children with and without dyslexia. Is it possible to relate the cognitive and neurophysiological dynamics supporting the performances in terms of the shape of the response time PDF? If the aggregate distributions of the two groups happen to be rescaled versions of each other, we hypothesize that the networks of perceptual, cognitive, and neurophysiological processes supporting the performances is organized in a qualitatively similar manner, as the only difference entailed by rescaling is time dilation or time contraction. By contrast, shape differences are symptomatic of more substantial dynamical changes. At minimum, the behavior of the BM model indicates that shifts in scaling exponents indicate connection strength changes in the process network supporting the performance. Scaling exponent shifts can also be symptomatic of more dramatic changes, such as a functional re-organization of the supporting perceptual, cognitive and neurophysiological systems, resembling a phase transition in nonlinear dynamical system.

Consider two PDFs, PDF_1 and PDF_2 , depicting data from two experimental cells. Mean rescaling is defined such that if:

$$PDF_2(c \cdot x) = \frac{1}{c} PDF_1(x) \quad (6)$$

where

$$c = \frac{mean_2}{mean_1} \quad (7)$$

then the two PDFs are rescaled, with one group simply being proportionally faster (or slower) than the other. Ordinarily, the rescaling of an analytically defined PDF is achieved via rescaling of the scale parameter of a distribution. A familiar example of a scale parameter is the SD or σ of the normal distribution:

$$N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8)$$

where the mean μ is a familiar example of a location parameter. Indeed, subjecting Eq. (8) to $\sigma \rightarrow c \cdot \sigma$ (and simultaneously $\mu \rightarrow c \cdot \mu$) produces a set of two PDFs related by Eqs. (6) and (7). Rescaling the variable, $x \rightarrow c \cdot x$, in PDF_2 will bring the two distribution back into one via $c \cdot PDF_2(c \cdot x) = PDF_1(x)$. On the other hand, for a lognormal distribution,

$$LN(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (9)$$

σ is a shape parameter and μ – or rather $\exp(\mu)$ – is a scale parameter. The lognormal is a heavy-tail distribution, which enjoys wide utility across many fields (Limpert, Stahel, & Abbt,

2001). It is used here as the front end of the lognormal-Pareto distribution. Just as the normal and stable distributions provide a framework for additive systems via the central limit theorem and the generalized central limit theorem, the lognormal – as mentioned above – offers a framework for multiplicative interactions (i.e., additive in the logarithmic domain, see Holden et al., 2009; Van Rooij et al., 2013).

Nonparametric rescaling. A simple nonparametric test of whether two distributions are rescaled versions of each other can be conducted as follows. Under the scaling assumption (6) and (7), we have

$$\ln PDF_2(c \cdot x) = \ln PDF_1(x) - \ln c \quad (12)$$

and in particular:

$$\overline{\ln PDF_2(c \cdot x)} = \overline{\ln PDF_1(x)} - \ln c \quad (13)$$

Consequently, the test of scaling between a control group (C) data and an experimental group (E) can be compared as follows:

- 1.) evaluate the ratio of the mean of C and the mean of E;
- 2.) evaluate the difference of the mean of $\ln C$ and mean of lognormal E;
- 3.) zero out the mean of log-transformed data of each group by subtracting the mean of the respective log-transformed data;
- 4.) perform a two-sample Kolmogorov-Smirnoff test on the zero-mean log-transformed data using a 5% significance level, where rejecting the null means rejecting the rescaling hypothesis.

As noted, an *exact* rescaling results when multiplying by the ratio of two distribution's means aligns their respective PDFs. *Statistical* rescaling results if multiplying by *any* constant other than a ratio of the means allows one PDF to approximate a second density function. While it is interesting to see exact self-similarity in an empirical system, statistical self-similarity is more common. For instance, coastlines are statistically self-similar, and their characteristic scaling exponents may change as a function of scale. Likewise, only $1/f^1$ noise is exactly self-similar in the sense that multiplying the time and amplitude axes by the same constant yields a statistically equivalent noise. However, scientists generally consider any spectral scaling exponent statistically greater than zero, and less than 3 to be a fractal noise. It is just that the time and amplitude axes must be rescaled by different proportions to yield identical self-affinity at a new scale.

To summarize, if two RT distributions from well controlled experiments can be transformed into each other via rescaling without affecting their shape parameters, then the cognitive and physiological networks supporting the performances are effectively the same. By contrast, substantive changes in the shape parameters is symptomatic of more substantial quantitative changes in systems dynamics, and may even signal an alternate functional organization.

DYSLEXIA MODELING STUDY

We now use the generalized inverse gamma and lognormal-Pareto distributions to characterize the differences between the dyslexic and control groups as a function of the tasks described in detail in Holden et al. (2014). The original article contrasted the performance of children with dyslexia and matched age-appropriate readers in four laboratory tasks with varied reading related requirements. In Holden et al., each participant's data was fit with an lognormal-Pareto and the obtained parameters were averaged across subjects. That analysis suggested little difference between the control and dyslexic groups in the flanker task, and the group differences

in word naming were consistent with a rescaling. However, the method relied on averaging parameters across individual fits, and was likely less sensitive to between-group differences than the omnibus fitting outlined in the simulation described earlier.

Here we adopt the procedure established in the earlier simulations to reveal additional information. We re-analyze the flanker, color naming, word naming, and arithmetic tasks (Holden et al., 2014). The tasks differed according to a loose rank order with respect to their differential emphasis on perceptual, linguistic, and higher-level cognitive demands. The flanker task emphasizes response inhibition and elementary perceptual-motor skills. A color-naming task emphasizes speaking aloud without reading. Speeded naming emphasizes reading aloud. An arithmetic task elicits high-level symbol-based cognitive activity without an explicit requirement for reading.

Once we fit the distributions with generalized inverse gamma and lognormal-Pareto we used a bootstrap procedure (Efron et al., 1994) to determine the CIs and distributions of the parameters. As in the simulations, we use two shape-related markers to gauge variability: The HW of the distribution and the exponent of the power-law tail of PDF. HW, as defined above, is illustrated in Figure 1. The product of HW and MPDF defines the area of the dashed box in Figure 1. For sufficiently large samples, this product typically lies in a narrow range centered around 0.8, as illustrated in Table 2 and approximates the fraction of PDF that is *not* in the tail.

Insert Table 2 about here

Table 2 displays the values and products of the half-width and the modal PDF. Again, the products approximate the fraction of the PDF that is *not* in the distribution's tail. The relatively smaller values for the group of children diagnosed with dyslexia indicate the respective distribution's tails occupy relatively more of the PDF's area.

Method

Participants and Procedure. The modeled data is from Holden, Greijn, van Rooij, Wijnants, and Bosman (2014). Twenty 6th grade Dutch students with dyslexia and 23 age-matched controls completed four perceptual and cognitive tasks that measured response time and errors as dependent measures. Each task administered a total of 560 trials. The present analyses include only data that passed the task-by-task inclusion criteria detailed in the Holden et al. method to preserve outcome comparability. On each flanker trial, the flanker task required participants to press one of two buttons indicating whether the center target arrow in a row of five arrows was pointing to the left or right. On each color naming trial, participants pronounced the color of a visually presented color patch as quickly and accurately as possible. Again, RT was measured as the interval between the presentation of the color patch and the onset of a participant's pronunciation. The experimenter coded accuracy manually during each participant's color naming session. On each naming trial, participants pronounced individually presented words as quickly as possible. RT was measured as the interval between stimulus word presentation and the onset of a pronunciation. The experimenter coded accuracy manually during each participant's naming session. On each arithmetic trial, participants pressed a "yes" button to indicate that a visually presented addition equation was correct (e.g., $2+2=4$), or a "no" button for incorrect sums (e.g., $3+4=2$). Half the trials depicted correct equations with answers being numbers less than ten, the other half were incorrect equations with answers being numbers less than ten.

Results

According to the BIC and KS goodness-of-fit statistics, in each task both power-law models provide better descriptions of the empirical control and dyslexic distributions than two illustrative RT models, the exGaussian and the Weibull distributions (see tables 4 & 5.). Each plot in Figure 4 allows for a visual comparison of control and dyslexia power-law models as a function of task.

Flanker. Plots A and B in Figure 4 depict the control and dyslexia data and fits for the generalized inverse gamma and lognormal-Pareto models. The distributions are very similar, and lie nearly atop one another when depicted on an axis that accommodate all tasks. Plots A and B of Figure 5 depict the bootstrap-resampled distribution of the generalized inverse gamma $\alpha\gamma + 1$ scaling exponents and the lognormal-Pareto $\alpha + 1$ scaling exponents for both the control and dyslexia groups. Both model PDFs reveal only subtle differences between the two groups. Indeed, the bootstrap analyses indicated overlapping parameter distributions, with different means. In this case, the RT performances of the children diagnosed with dyslexia can be characterized as a near exact mean rescaling of the distribution of control performances. Similarly, the rescaling hypothesis could not be rejected by the nonparametric rescaling test. From a dynamical perspective, the rescaling can be interpreted as indicating a similar dynamical organization supporting both group's performances. However, the performances of the group with dyslexia are slightly dilated in time, relative to the control group.

Color Naming. The color naming results are presented in plots C and D of Figure 4. Relative to flanker performance, the shapes of control and dyslexia RT distributions diverged. The bootstrap analyses revealed that while the faster front ends of the distributions remained comparable, the scaling exponents describing the distribution's tails were reliably different. Nevertheless, the plots indicate the shape differences to be relatively incremental. Thus, the PDFs are inconsistent with an exact mean rescaling, and only marginally approximate a more general statistical rescaling.

Word Naming. The empirical word naming histograms and model fits are presented in plots E and F of Figure 4. The bootstrap distributions, depicted in plots E and F of Figure 5, revealed that relatively distinct scaling exponents characterize the control and dyslexia word naming performances. Both models indicate the dyslexia distribution's tail is more skewed than the control group. Moreover, the HW of the control and dyslexia distributions is smaller than color naming indicating the tails of both word naming distributions occupy an increased portion of their respective density functions. From the perspective of the task progression, the shapes of word naming densities changed in a similar manner, but the changes in the dyslexia group were exaggerated relative to the control group.

Arithmetic. The arithmetic results are depicted in plots G and H of Figure 4. Both the generalized inverse gamma and lognormal-Pareto bootstrap contrasts of the arithmetic distributions — plots G and H in Figure 5 — also revealed distinct control and dyslexia scaling exponents. The arithmetic task produced the smallest scaling exponents, and thus the most prominent power-law behavior among the four tasks.

Insert Table 3 about here

Table 3 depicts the mean scaling exponents and confidence intervals derived from the fits of the generalized inverse gamma and lognormal-Pareto models. The scaling exponents are depicted as a function of task and group. Smaller scaling exponents correspond to distributions with more dramatically skewed right tails. Both models indicate the word naming control

distribution has the largest scaling exponent. The word naming scaling exponent for the group of children diagnosed with dyslexia is comparable to the scaling exponents from several of the other tasks.

The mean scaling exponents and confidence intervals are derived from the fits of the generalized inverse gamma and lognormal-Pareto models. The scaling exponents are depicted as a function of task and group. Smaller scaling exponents correspond to distributions with more dramatically skewed right tails. Both models indicate the word naming control distribution has the largest scaling exponent. The word naming scaling exponent for the group of children diagnosed with dyslexia is comparable to the scaling exponents from several of the other tasks.

Insert Figure 4 about here

The two plots in each row of Figure 4 both depict the empirical control and dyslexia data sets as darker and lighter histograms, respectively. The Flanker data is in the top row (A & B), the color naming is in the second row (C & D), word naming appears in the third row (E & F) and arithmetic appears in the fourth row (G & H). The left-hand plots illustrate the generalized inverse gamma fits of the two groups as darker and lighter lines, respectively. Likewise, the right-hand plots illustrate the lognormal-Pareto fits of each group as darker and lighter grey lines. Both models achieved reasonably successful fits of the data sets, relative to fits of other models that are commonly used to represent response times (e.g., see Tables 4 & 5).

Insert Figure 5 about here

The two plots in each row of Figure 5 both depict the bootstrapped scaling exponents for the control and dyslexia data sets as darker and lighter histograms, respectively. Bootstrapped flanker exponents in the top row (A & B), the color naming exponents are in the second row (C & D), word naming exponents appear in the third row (E & F) and arithmetic exponents appears in the fourth row (G & H). The left-hand plots illustrate scaling exponents estimated with the generalized inverse gamma fits and darker and lighter histograms, respectively. Likewise, the right-hand plots illustrate scaling exponents derived from lognormal-Pareto fits as darker and lighter gray histograms. Both models arrived at quantitatively similar depictions of the relative magnitudes of the scaling exponents in each task by group cell. This suggests a relatively consistent underlying pattern of empirical differences.

Insert Table 4 about here

Table 4 presents the Bayesian information criterion (BIC) fitting outcomes of fits for the group data, using the two versions of the generalized inverse gamma (GIGa), the lognormal-Pareto (LNP), the exGaussian, and the Weibull. The BIC penalizes models with more free parameters. The bolded values are the best fits. This indicates the lognormal-Pareto was the most successful model.

Insert Table 5 about here

Table 5 presents the results of Kolmogorov–Smirnov (KS) difference tests for several model fits of the group data. Models included two versions of the generalized inverse gamma

(GIGa), the lognormal-Pareto (LNP), the exGaussian, and the Weibull. The bolded values are the best fits. The KS tests of the maximum difference between the model and empirical distributions and also indicated the lognormal-Pareto as the most successful model. In contrast to the BIC, the KS test does not impose a penalty for the number of model parameters.

Discussion

The shapes of the control and dyslexia response time distributions diverged as a function of task. With respect to the two groups, the flanker distributions were very similar, and related by a mean rescaling. The color naming distributions shared similar half-widths. Their shapes were statistically distinct, as indicated by their scaling exponents which differed by between $\frac{1}{2}$ to 1 unit, as estimated by the lognormal-Pareto and generalized inverse gamma models, respectively. The control and dyslexic word naming distributions maintained similar halfwidths, but somewhat more distinct shapes than the color naming distributions. Their scaling exponents differed by between 1 and 3 units, according to the lognormal-Pareto and generalized inverse gamma fits. Finally, bootstrap contrasts of two group's arithmetic distributions revealed distinct half-widths and distinct shapes. The arithmetic scaling exponents differed by about 1 to 1.5 units, according to the lognormal-Pareto and generalized inverse gamma fits. The absolute difference between the arithmetic exponents is relatively small numerically, but relative to the other tasks, both are closest to the minimum scaling exponent values that are permitted by these statistical models.

The word naming task displayed the greatest between-group shape differences. However, the average word naming scaling exponents were larger than those of the remaining tasks, and were therefore the most compact and least skewed distributions. The flanker, color naming, and arithmetic tasks yielded progressively more robust scaling behavior. Thus, in terms of the tasks, word naming yields the most constrained power-law behavior and arithmetic the most salient power-law behavior.

A potential inconsistency with the aforementioned task continuum is that reading printed text is arguably the most difficult of the four tasks, especially for children diagnosed with dyslexia. If larger scaling exponents are associated with increased skill, then why does word naming tend to return such large scaling exponents? It is important to remember that school children receive extensive reading and literacy instruction. Most schools begin implementing literacy-oriented curricula by the first grade. Thus, even 6th graders with weaker reading skills have nevertheless learned task relevant skills. By contrast, the remaining tasks are relatively novel for both groups of children. For instance, the button-press tasks impose time pressure and arbitrary mappings between their binary responses and the children's hands (cf. Holden et al., 2014). It is the control participant's word naming distribution that is truly unique among the Task by Group sub-distributions or experimental cells. Unsurprisingly, it is consistent with a skilled performance. It is notable that the dyslexic children's word naming distributions are reasonably similar to the other examples of relatively less-skilled performances indexed by the remaining experimental cells.

Historically, debate in the dyslexia literature questioned if dyslexia was most productively viewed as a qualitative or quantitative impairment. That is, do the reading skills of children diagnosed with dyslexia and their age-matched controls differ by degree, or in kind? The in-kind view interprets reading difficulties as the result of an absence or dysfunction of a reading-dedicated cognitive component. The historical emphasis on contrasts of mean naming times between controls and participants with dyslexia reinforced this interpretation. For example, simple location shifts imply a compensatory processing delay. Indeed, the mean response times in the word naming task differ the most between the dyslexic and control groups.

However, the modeling reveals that the mean differences are driven by relatively consistent but incremental distribution shape changes, both across and within tasks. This outcome suggests relatively subtle between-group differences in low-level perceptual, cognitive, or neurophysiological dynamics that are progressively amplified as a function of task difficulty and relative skill (Holden et al., 2014). From this perspective, the word naming task reveals the largest group differences because it requires much more extensive skills than any of the remaining tasks. The shape of control group's word naming distributions is most consistent with the relatively constrained, well-organized dynamics of skilled performance. The shapes of the remaining distributions, including the dyslexic participant's word naming distributions are more consistent with the less constrained dynamics of less skilled performance (e.g., see Holden et al., 2009). As noted, historically the term dyslexia meant a qualitative difference. Due to the difficulty of distinguishing between those with and those without a reading disorder, contemporary researchers advocate for alternative terms, such as good versus poor decoders, to emphasize the continuum in relative reading skills illustrated by the two participant groups (A. Bosman, personal communication, April 22, 2019).

General Discussion

The power-law response time models revealed a relatively consistent pattern of decreasing scaling exponents, both within tasks, in contrasts between groups, and across tasks, as a function of increasing task complexity. Thus, progressively *decreasing* scaling exponents track a loose continuum of progressively *increasing* task difficulty, within which the control group expressed more stable performances. This outcome is notable because the incremental nature of the within-task shape changes tends to mirror between-task shape differences.

Cognitive manipulations that exclusively produce mean shifts in response time distributions are exceedingly rare (Holden et al., 2009). From a dynamical perspective, distribution rescaling illustrates a least-common denominator in the boundary between the presence and absence of a cognitive effect. Elsewhere, we explained how rescaling likely results from the mismatch between *absolute* clock time and the change-based *relative* time arising from the rate-limiting, often proportional, physiological transactions that are characteristic of biological systems (Holden, 2013; Holden, Ma, & Serota, 2013).

On the linear axis of response time, proportional dynamics appear as a lognormal-shaped distribution. More exaggerated power-law behavior results from the emergence of feedback dynamics rooted in less stable or less constrained cognitive and bodily dynamics. This proposal yields concrete behavioral hypotheses, despite the fact that specific biological or physiological markers are rarely observed directly in response time tasks.

Consider the classic word frequency effect in lexical decision: Faster and more accurate response times to commonly encountered words, as compared to less common words. Conventional connectionist narratives attributed the frequency advantage to stronger relationships among a word's orthographic, phonological, and semantic dimensions (e.g., McClelland & Rumelhart, 1981). The stronger associative connections were posited to arise as a consequence of learning, resulting from encountering and using common words more often. The Bouchard-Mezard network dynamics predict the relative strengths of the dynamical relationships supporting a performance influence response time scaling exponents. As such, network principles predict smaller response time scaling exponents to low versus high frequency words. This hypothesis was recently confirmed in large standard lexical decision task that manipulated word frequency using the lognormal-Pareto model (Amon & Holden, 2019). The same outcome

is mirrored in the scaling exponent differences revealed by the control versus dyslexia group contrasts. Apparently, stronger connection strengths support more skilled reading, and weaker connection strengths support less skilled reading.

Likewise, studies exploring related hypotheses revealed similar results. For instance, Cooper and Shepard's classic letter rotation task asks participants to distinguish normal letters from mirror reversed versions when presented at various orientations. Statistical rescaling results from small deviations up to about 60 degrees from their upright orientations. Larger deviations exceeding 120 decrease scaling exponents, indicating performances that are supported by weaker coordinative dynamics (van Rooij, Nash, Rajaraman & Holden, 2013). A related approach revealed subtle differences in the eye-movement latencies that distinguished between patients diagnosed with prodromal and mild Alzheimer's disease from each other, and appropriate controls (Holden et al., 2018). In the future, we hope to extend our approach to additional clinical contexts, such as studies of the efficacy of ADHD medications (e.g., Epstein et al., 2011).

Historically, cognitive scientists are hesitant to link physiological dynamics to cognitive dynamics unless reductive or causal relations are established as a basis for such discussion. In this article, the entry point for the claim of a link is renormalization group theory (e.g., Bruce & Wallace, 1989). As one zooms out from micro to macro time-scales, self-similarity and interdependence are implicated by the qualitatively similar PDF's observed across various time-scales. In this view, neurophysiological variability is *coarse-grained* in behavioral measurements, and is thus echoed in measures of behavioral variability.

This coarse-graining hypothesis is generally consistent with independently conducted physiological studies. For instance, heavy tailed lognormal distributions are symptomatic of autocatalytic growth in many biological processes (Koch, 1966). They indicate multiplicative amplification, and are evident across multiple physiological scales. Thus, it is unsurprising that lognormal distributions are characteristic of many neurophysiological properties of the brain (Bassingthwaite et al., 1994). They characterize the firing rates and bursting behavior of individual neurons. They depict the range of the relative synaptic connection strengths among neurons, and the durations of temporal synchrony events that emerge in populations of spiking neurons (Buzsáki & Mizuseki, 2014). Eye-movements typically unfold on slightly slower time-scales than the aforementioned neurophysiological dynamics, but they do typically unfold on faster timescales than the behavioral measurements that are representative of response time tasks. Eye-movement latency distributions, themselves are heavy-tailed, and typically correspond to the lognormal (Holden et al., 2018). So, when lognormal and power-law distributions are found to characterize response times (e.g., Holden et al., 2009; 2012), the most plausible working hypothesis is that response times inherit contingencies from faster-time scale neurophysiological events because their variability patterns are constrained by those supporting processes. The rescaling hypothesis articulates this perspective, as time contraction and dilatation are self-affine transformations — not qualitative changes. By contrast, a change in the characteristic scaling exponent suggests aspects of the micro-macro control hierarchy have somehow changed.

REFERENCES

- Amazeen, P. G., Schmidt, R. C., & Turvey, M. T. (1995). Frequency detuning of the phase entrainment dynamics of visually coupled rhythmic movements. *Biological Cybernetics*, 72(6), 511-518.
- Amon, M. J., & Holden, J. G. (2019). The mismatch of intrinsic fluctuations and the static

- assumptions of linear statistics. *Review of Philosophy and Psychology*, <https://doi.org/10.1007/s13164-018-0428-x>.
- Amon, M. J., Pavlov, O. C., & Holden, J. G. (2018). Synchronization and fractal scaling as foundations for cognitive control. *Cognitive Systems Research*, *50*, 155-179.
- Ancelin, M. L., Artero, S., Portet, F., Dupuy, A. M., Touchon, J., & Ritchie, K. (2006). Non-degenerative mild cognitive impairment in elderly people and use of anticholinergic drugs: longitudinal cohort study. *BMJ*, *332*, 455-459.
- Bassingthwaighe, J. B., Liebovitch, L. S., & West, B. J. (2013). *Fractal physiology*. New York: Springer.
- Bouchaud, J. P., & Mézard, M. (2000). Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, *282*, 536-545.
- Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., ... & Brammer, M. (2001). Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Human brain mapping*, *12*(2), 61-78.
- Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford, UK: Oxford University Press.
- Buzsáki, G., & Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*, *15*(4), 264.
- Castañeda, T. R., de Prado, B. M., Prieto, D., & Mora, F. (2004). Circadian rhythms of dopamine, glutamate and GABA in the striatum and nucleus accumbens of the awake rat: modulation by light. *Journal of pineal research*, *36*(3), 177-185.
- Cavanaugh, J. T., Kelty-Stephen, D. G., & Stergiou, N. (2017). Multifractality, interactivity, and the adaptive capacity of the human movement system: A perspective for advancing the conceptual basis of neurologic physical therapy. *Journal of neurologic physical therapy*, *41*(4), 245-251.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: CRC Press.
- Epstein, J. N., Brinkman, W. B., Froehlich, T., Langberg, J. M., Narad, M. E., Antonini, T. N., & Altaye, M. (2011). Effects of stimulant medication, incentives, and event rate on reaction time variability in children with ADHD. *Neuropsychopharmacology*, *36*, 1060-1072.
- Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134-140.
- Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*, 403-408.
- Farmer, J. D., & Geanakoplos, J. (2008, May). *Power laws in economics and elsewhere*. Retrieved January 1, 2016 from <http://www.elautomataeconomico.com.ar/download/papers/Farmer-powerlaw3.pdf>.
- Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews neuroscience*, *8*(9), 700.
- Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological review*, *108*(1), 33.
- Gonzales, L. M., Hessler, E. E., & Amazeen, P. G. (2010). Perceptual constraints on frequency ratio performance in motor-respiratory coordination. *Ecological Psychology*, *22*(1), 1-23.
- He, B. J. (2011). Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *Journal of Neuroscience*, *31*(39), 13786-13795.

- Holden, J. G. (2005). Gauging the fractal dimension of response times from cognitive tasks. In M. A. Riley & G. C. Van Orden (Eds.), *Contemporary Nonlinear Methods for Behavioral Scientists: A webbook tutorial* (pp. 267-318). Retrieved April 8, 2005, from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp>
- Holden, J. G. (2013). Cognitive effects as distribution rescaling. *Ecological Psychology, 25*, 256-266.
- Holden, J. G., Cosnard, A., Laurens, B., Asselineau, J., Biotti, D., Cubizolle, S. et al. (2018). Prodromal Alzheimer's Disease Demonstrates Increased Errors at a Simple and Automated Anti-Saccade Task. *Journal of Alzheimer's disease, 65*(4), 1209-1223.
- Holden, J. G., & Rajaraman, S. (2012). The self-organization of a spoken word. *Frontiers in Psychology, 3*, 209.
- Holden, J. G., Greijn, L. T., van Rooij, M. M., Wijnants, M. L., & Bosman, A. M. (2014). Dyslexic and skilled reading dynamics are self-similar. *Annals of Dyslexia, 64*, 202-221.
- Holden, J. G., Ma, T., & Serota, R. A. (2013). Change is time: A comment on "Physiologic time: A hypothesis". *Physics of Life Reviews, 10*, 231.
- Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review, 116*, 318-342.
- Ihlen, E. A. (2013). The influence of power law distributions on long-range trial dependency of response times. *Journal of Mathematical Psychology, 57*, 215-224.
- Izhikevich, E. M. (2007). *Dynamical systems in neuroscience*. Cambridge, MA: MIT press.
- Jensen, H. J. (1998). *Self-organized criticality: Emergent complex behavior in physical and biological systems*. Cambridge, UK: Cambridge University Press.
- Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review, 121*(1), 1-32.
- Kello, C. T., Anderson, G. G., Holden, J. G., & Van Orden, G. C. (2008). The pervasiveness of $1/f$ scaling in speech reflects the metastable basis of cognition. *Cognitive Science, 32*(7), 1217-1231.
- Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General, 136*(4), 551-568.
- Kelso, J. S. (1997). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT press.
- Kelso, J. S. (2012). Multistability and metastability: Understanding dynamic coordination in the brain. *Phil. Trans. R. Soc. B, 367*(1591), 906-918.
- Kugler, P. N., & Turvey, M. T. (2015). *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Lawrence Erlbaum.
- Lam, H., Blanchet, J., Burch, D., & Bazant, M. Z. (2011). Corrections to the central limit theorem for heavy-tailed probability densities. *Journal of Theoretical Probability, 24*, 895-927.
- Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal Distributions across the Sciences: Keys and Clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *AIBS Bulletin, 51*, 341-352.
- Linkenkaer-Hansen K, Nikouline VV, Palva JM, Ilmoniemi RJ (2001). Long-range temporal correlations and scaling behavior in human brain oscillations. *Journal of Neuroscience, 21*(4):1370-1377.

- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford, UK: Oxford University Press.
- Ma, T., & Serota, R. A. (2014). A model for stock returns and volatility. *Physica A: Statistical Mechanics and its Applications*, 398, 89-115.
- Ma, T., Holden, J. G., & Serota, R. A. (2013). Distribution of wealth in a network model of the economy. *Physica A: Statistical Mechanics and its Applications*, 392, 2434-2441.
- Ma, T., Holden, J. G., & Serota, R. A. (2016). Distribution of human response times. *Complexity*, 21, 61-69.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375-407.
- Moscoso Del Prado Martin F (2008). *A theory of reaction time distributions*. Retrieved January 1, 2016 from <http://cogprints.org/6310/1/recinormal.pdf>
- Nolan, J. P. (2005). *Stable distributions: Models for Heavy Tailed Data*. Retrieved January 1, 2016 from <http://fs2.american.edu/jpnolan/www/stable/chap1.pdf>.
- Osmon, D. C., Kazakov, D., Santos, O., & Kassel, M. T. (2018). Non-Gaussian Distributional Analyses of Reaction Times (RT): Improvements that Increase Efficacy of RT Tasks for Describing Cognitive Processes. *Neuropsychology review*, 28(3) 359-376.
- Riley, M. A., & Holden, J. G. (2012). Dynamics of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(6), 593-606.
- Seuront, L. (2009). *Fractals and multifractals in ecology and aquatic science*. Boca Raton, FL: CRC Press.
- Sigman, M., Etchemendy, P., Fernandez Slezak, D., & Cecchi, G. A. (2010). Response time distributions in rapid chess: A large-scale decision making experiment. *Frontiers in Neuroscience*, 4, 60.
- Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 744-774.
- 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(3), 300-315.
- Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132, 331-350.
- Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human cognition and $1/f$ scaling. *Journal of Experimental Psychology: General*, 134, 117-123.
- Van Orden, G., Hollis, G., & Wallot, S. (2012). The blue-collar brain. *Frontiers in Physiology*, 3, 207.
- van Rooij, M. M., Nash, B. A., Rajaraman, S., & Holden, J. G. (2013). A fractal approach to dynamic inference and distribution analysis. *Frontiers in Physiology*, 4, 1.
- West, B. J., Deering, B., & Deering, W. D. (1995). *The lure of modern science: Fractal thinking*. Hackensack, NJ: World Scientific.
- Wijnants, M. L. (2014). A review of theoretical perspectives in cognitive science on the presence of scaling in coordinated physiological and cognitive processes. *Journal of Nonlinear Dynamics*, 2014.
- Wirz-Justice, A. (1987). Circadian rhythms in mammalian neurotransmitter receptors. *Progress in Neurobiology*, 29(3), 219-259.

Table 1: Comparisons of Actual and Simulated Parameter Estimates

Task: Arithmetic			Control			Dyslexic	
Data Type	Parameter	Lower 95% CI	M (SD)	Upper 95% CI	Lower 95% CI	M (SD)	Upper 95% CI
Group	$\alpha\gamma+1$	4.47	4.66 (.16)	5.11	2.95	3.11 (.09)	3.29
Simulated G	$\alpha\gamma+1$	4.8	4.87 (.2)	5.14	3.06	3.11 (.05)	3.24
Group	HW	0.54	.56 (.02)	0.6	0.38	.44 (.03)	0.49
Simulated G	HW	0.56	.57 (.01)	0.59	0.43	.45 (.01)	0.47
Individual	$\alpha\gamma+1$	11.51	5.38 (2.24)	3.14	2.1	3.8 (1.24)	6.88
Simulated I	$\alpha\gamma+1$	[2.4]	4.94 (1.56)	[7.76]	[1.9]	3.13 (.71)	[4.27] Rng
Individual	HW	0.36	0.46 (.09)	0.63	0.38	.44 (.03)	0.49
Simulated I	HW	0.3	.52 (.08)	0.6	0.14	.41 (.11)	0.72

Table 2: Half-Width and Modal PDF values.

	HW	MPDF	HW \times MPDF
Arithmetic Control	0.679	1.215	0.825
Arithmetic Dyslexia	0.742	1.002	0.743
Flanker Control	0.235	3.516	0.825
Flanker Dyslexia	0.236	3.464	0.818

Table 3: Half-Widths and Average Scaling Exponents by Task and Group

Task	Condition	HW	Power law Model					
			GiGa $\alpha\gamma+1$			LNP $\alpha+1$		
			5%	M	95%	5%	M	95%
Flanker	Control	0.45	4.54	4.90	5.29	4.06	4.23	4.58
	Dyslexia	0.46	4.35	4.66	4.91	3.64	3.91	4.22
Color Naming	Control	0.48	4.83	5.03	5.21	4.52	4.60	4.81
	Dyslexia	0.41	4.18	4.40	4.63	3.95	4.01	4.17
Word Naming	Control	0.35	6.45	7.69	8.21	4.84	5.40	6.13
	Dyslexia	0.32	4.29	4.48	4.99	4.23	4.47	4.82
Arithmetic	Control	0.58	4.47	4.66	5.11	3.72	3.87	3.98
	Dyslexia	0.39	2.95	3.11	3.29	2.78	2.85	2.99

Table 4: Bayesian information criterion (BIC) fitting results

Data type	GIGa	LNP	GIGa3	exGaussian	Weibull
Flank Control	-8945.49	-10930.81	-7353.03	-5783.67	-2176.32
Flank Dyslexia	-6976.41	-8254.16	-6380.29	-6263.99	54.84
Color Control	2968.64	1890.10	33131.87	3355.69	6670.21
Color Dyslexia	8552.18	-415.58	13282.01	12549.15	15921.20
Name Control	-2117.84	-8642.09	-1389.72	-1142.68	538.68
Name Dyslexia	-899.043	-7601.006	517.740	4483.709	6769.26
Arith. Control	13348.55	9685.56	13791.30	14685.69	18881.39
Arith. Dyslexia	13692.73	10263.34	15363.81	18702.14	16148.29

Table 5. Kolmogorov–Smirnov (KS) difference fitting results.

Data type	GIGa	LNP	GIGa3	exGaussian	weibull3
Flank Control	0.02	0.008	0.025	0.039	0.072
Flank Dyslexia	0.034	0.015	0.039	0.055	0.091
Color Control	0.029	0.019	0.032	0.032	0.072
Color Dyslexia	0.025	0.017	0.03	0.047	0.076
Name Control	0.026	0.012	0.029	0.029	0.066
Name Dyslexia	0.035	0.014	0.040	0.057	0.090
Arith. Control	0.022	0.015	0.025	0.032	0.067
Arith. Dyslexia	0.026	0.019	0.028	0.068	0.056

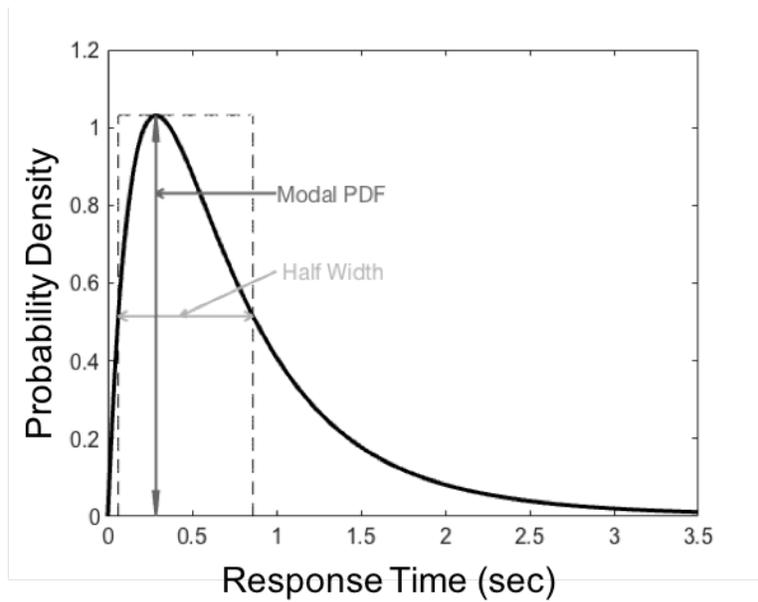
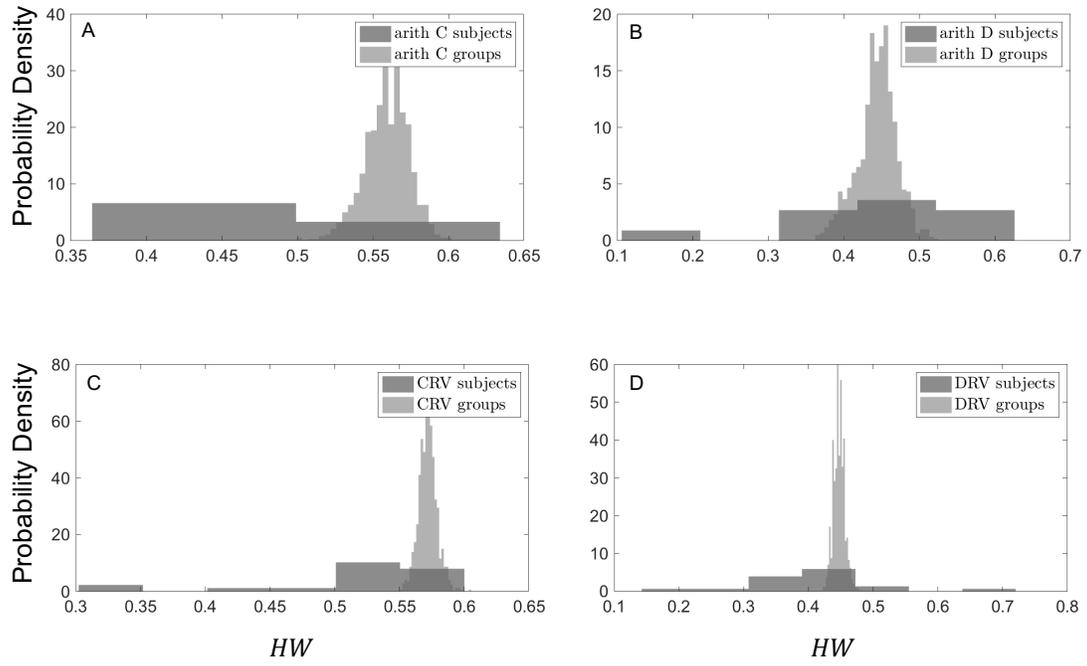
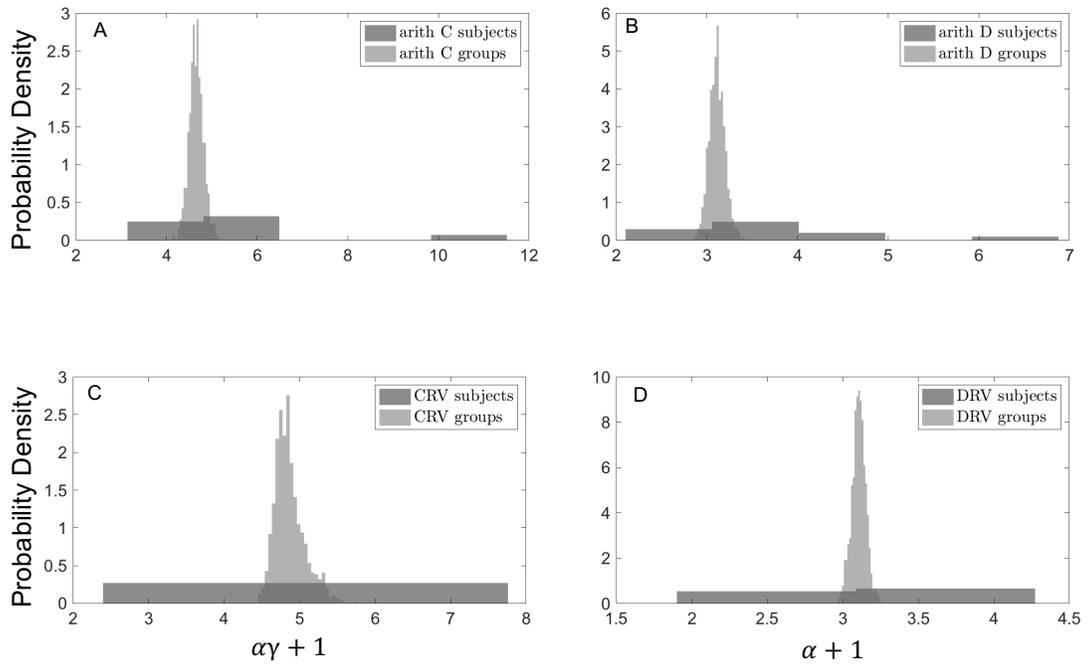


Fig. 1

**Fig. 2**

**Fig. 3**

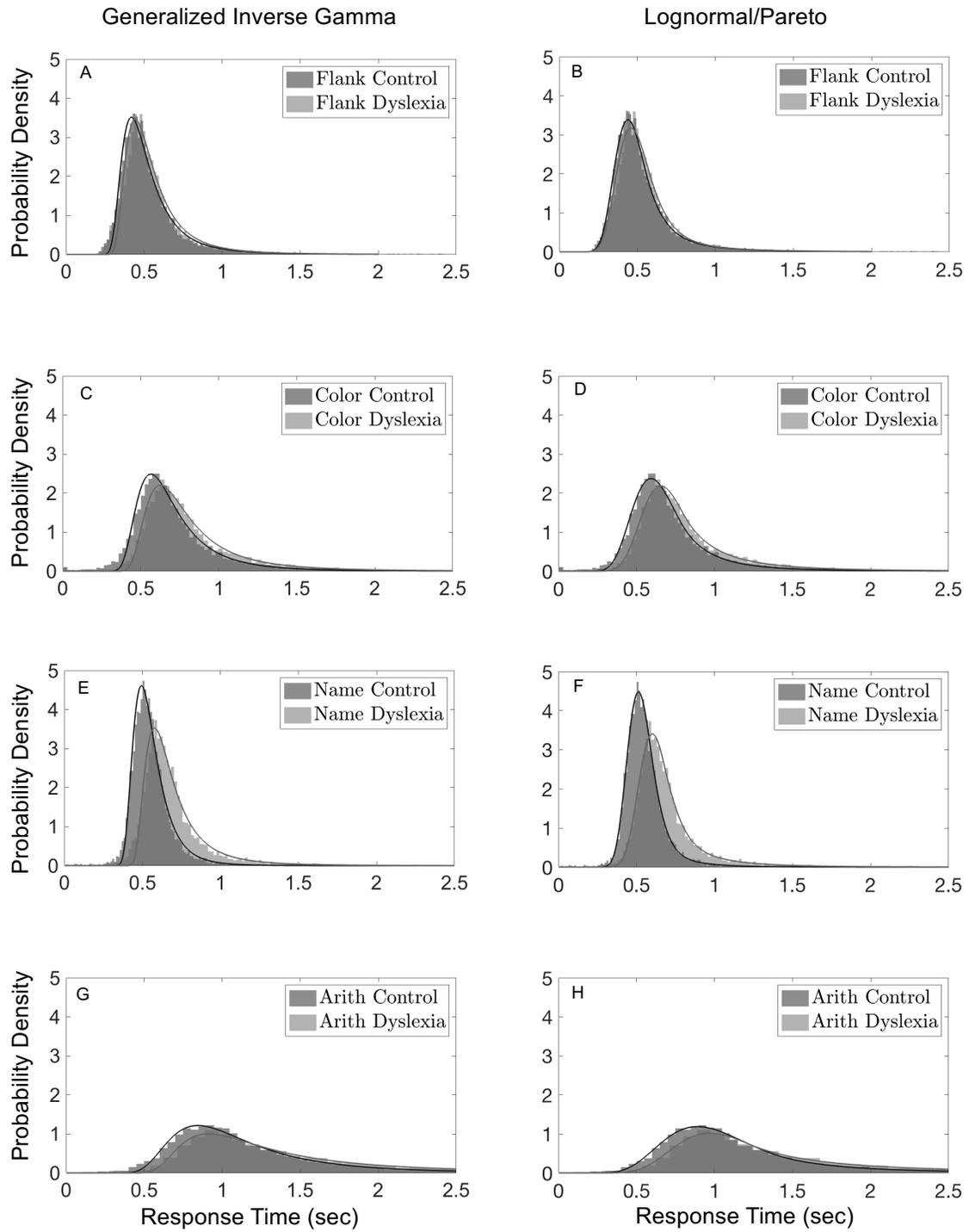


Fig 4

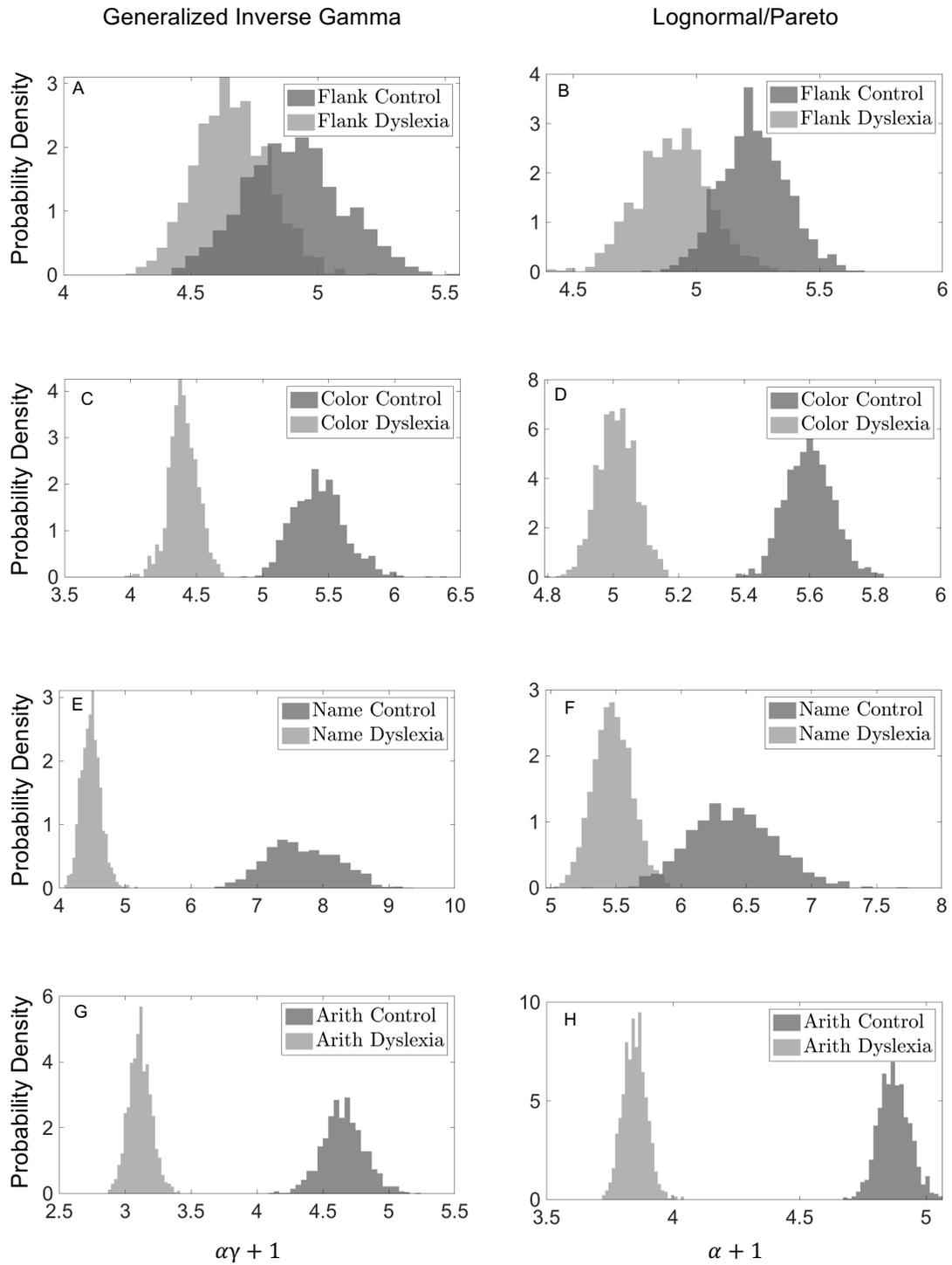


Fig. 5