Random local temporal structure of category fluency responses

David J. Meyer, Jason Messer, Tanya Singh, Peter J. Thomas, Wojbor A. Woyczynski, Jeffrey Kaye & Alan J. Lerner

Journal of Computational Neuroscience

ISSN 0929-5313
Volume 32
Number 2

J Comput Neurosci (2012) 32:213-231
DOI 10.1007/s10827-011-0349-5

Springer

J Comput Neurosci (2012) 32:213–231
DOI 10.1007/s10827-011-0349-5

Random local temporal structure of category fluency responses

David J. Meyer · Jason Messer · Tanya Singh · Peter J. Thomas · Wojbor A. Woyczynski · Jeffrey Kaye · Alan J. Lerner

Received: 23 September 2010 / Revised: 6 June 2011 / Accepted: 8 June 2011 / Published online: 8 July 2011

Abstract The Category Fluency Test (CFT) provides a sensitive measurement of cognitive capabilities in humans related to retrieval from semantic memory. In particular, it is widely used to assess progress of cognitive impairment in patients with dementia. Previous research shows that, in the first approximation, the intensity of tested individuals' responses within a standard 60-s test period decays exponentially with time, with faster decay rates for more cognitively impaired patients. Such decay rate can then be viewed as a global (macro) diagnostic parameter of each test. In the present paper we focus on the statistical properties of the properly de-trended time intervals between consecutive responses (inter-call times) in the Category Fluency Test. In a sense, those properties reflect the local (micro) structure of the response generation process. We find that a good approximation for the distribution of the de-trended inter-call times is provided by the Weibull Distribution, a probability distribution that appears naturally in this context as a distribution of a minimum of independent random quantities and is the standard tool in industrial reliability theory. This insight leads us to a new interpretation of the concept of "navigating a semantic space" via patient responses.

Keywords Category Fluency Test · Semantic memory · Cognitive impairment · Alzheimer's disease · Statistical temporal structure · Weibull distribution · Inter response times

Action Editor: P. Dayan

D. J. Meyer · J. Messer · T. Singh
Case Western Reserve University, 11900 Euclid Avenue, Cleveland, OH 44106, USA

P. J. Thomas
Departments of Mathematics, Biology and Cognitive Science, Case Western Reserve University, 11900 Euclid Avenue, Cleveland, OH 44106, USA

P. J. Thomas
Department of Neuroscience, Oberlin College, Oberlin, OH 44074, USA

W. A. Woyczynski (✉)
Department of Statistics and Center for Stochastic and Chaotic Processes in Science and Technology, Case Western Reserve University, 11900 Euclid Avenue, Cleveland, OH 44106, USA
e-mail: waw@case.edu

J. Kaye
Departments of Neurology and Biomedical Engineering, Oregon Health and Science University, Portland, OR 97239, USA

J. Kaye
Portland Veterans Affairs Medical Center, Portland, OR 97239, USA

A. J. Lerner
Department of Neurology, Case Medical Center, Case Western Reserve University, 11900 Euclid Avenue, Cleveland, OH 44106, USA

1 Introduction

1.1 Category Fluency Tests (CFT)

The Category Fluency Test (CFT) is a form of controlled oral word association used by Neuropsychologists or Cognitive Scientists to sample phonemic or semantic domains. One of the commonest category fluency tests is the one minute animal naming task, wherein a subject is asked to name as many animals as possible in the allotted time. Animal naming is well developed for use in children and adults, routinely applied, and is sensitive to early cognitive changes in Mild Cognitive Impairment and Alzheimer's disease (AD); it has been adapted as part part of the Neuropsychology Battery of the Uniform Data Set designed by the National Institute of Health, see, e.g., Caramelli et al. (2007), Gomez and White (2006), Kramer et al. (2006), Canning et al. (2004), Morris et al. (2006), Sauzeon et al. (2004), Tombaugh et al. (1999), Diaz et al. (2003) and Weiner et al. (2008).

The basic test metric of category fluency tests such as one minute animal naming is the total number of items (N60), although analyses of content relatedness and number per unit time (e.g., number named per 15-s epoch), have also been studied, see, Rohrer et al. (1995), Troyer et al. (1997) and Lerner et al. (2009).

Rohrer et al. (1995) elegantly demonstrated exponential decay in the intensity of tested individuals' responses within a standard 60-s test with faster decay rates for more cognitively impaired patients. Such decay rates can be viewed as a global (macro) diagnostic parameter of each test.

In the present paper we focus on the statistical properties of the properly de-trended times between consecutive responses, dubbed here "inter-call (iC) times", in the one minute animal naming task, whose properties reflect the local (micro) structure of the response generation process.

The first surprising observation is that, after taking into account the exponential trend, the inter-arrival times' distribution is not exponential thus eliminating the stationary Poisson Markov hypothesis for the de-trended response process. Instead, we found out that a better approximation is provided by the Weibull Distribution, a probability distribution that appears naturally in this context as a distribution of the minimum of independent random quantities and is a standard tool in industrial reliability theory (Weibull 1951). This insight leads us to a new interpretation of the concept of "navigating a semantic space" via subject responses. More importantly, we found significant differences between the random structure of the times between responses we interpret formally as "within the same category" of animals and those corresponding to "category switching". In addition, we report here on differences between younger adults and older adults.

The problem of the statistical distributions of what has been often called in mathematical psychology inter-response times (IRT) has been studied for a long time going back to the pioneering and influential 1944 efforts by Bousfield and Sedgwick (1944), although not necessarily in the specific context of the Category Fluency Tests. We will discuss the relationship of our work with the previous studies in more detail in Section 6.

1.2 The recall sequence

For each tested individual, the times of recall of (correct) consecutive items form a sequence

$$0 = t_0 < t_1 < t_2 < t_3 < \cdots < t_{N_{60}} \leq 60, \tag{1.1}$$

which has a random structure varying from individual to individual, and for tests applied to the same individual at different times. Rohrer et al. (1995) argued that, on the average, during the 60-s test the number of responses, say, per 10-s bin, decreases exponentially as the semantic space of each individual is being exhausted. We accept their argument, call this phenomenon *exponential exhaustion* and take it as a starting point of our work in which we will investigate statistical properties of the random fluctuations of the inter-call (iC) time interval sequence,

$$\delta t_1 = t_1 - t_0,\ \delta t_2 = t_2 - t_1,\ \ldots\ , \delta t_{N_{60}} = t_{N_{60}} - t_{N_{60}-1}. \tag{1.2}$$

For the purposes of this paper, the hypothesis of Rohrer et al., can be rephrased as a statement that the cumulative number of recalls $R(t)$, by the time $t, 0 < t < 60$, (measured in seconds) is approximately

$$R(t) \approx N_\infty (1 - e^{-t/\tau}), \tag{1.3}$$

with N_∞ representing the "total (asymptotic) number of items the individual could recall given infinite time", and the constant τ representing what is called in Rohrer et al. (1995) the *mean latency* of the recall process. The above quotation marks were inserted advisedly as N_∞ has nothing to do with the actual individual's supply of the names in a given semantic category; it is just an artificial parameter in the exponential model. Observe that the derivative $R'(t) = N_\infty e^{-t/\tau}/\tau$ is exactly Rohrer's function $r(t)$, which represents the rate at which the growth of the cumulative number of recalls slows down with time.

In an analysis of concrete experimental data, the parameters N_∞ and τ must be estimated to provide the best fit between the theoretical cumulative count function (1.3), and the empirical cumulative count function,

$$R_e(t) = \sum_{k=1}^{N} U(t - t_k), \tag{1.4}$$

where $U(t)$ is the unit step function equal to 0, for $t < 0$, and 1, for $t \geq 0$. $R_e(t)$ jumps up by 1 each time an item is called.

The primary goal of this paper is to provide a more detailed analysis of the statistical structure of random inter-call sequence $\delta t_1, \delta t_2, \ldots, \delta t_{N_{60}}$, defined in Eq. (1.2), after the above exponential "exhaustion" trend has been removed. It seems evident that this structure is dictated by the underlying dynamics of the cognitive process of retrieval from memory of consecutive items named by the individual during the test.

1.3 The structure of the paper

In Section 2 we describe our data set which comprises tests for subpopulations of young adults (YA), and older adults (OA), and provide estimation of the global parameters, N_∞, and τ, needed in formula (1.3). Then we compare the distributions of the three global parameters, N_{60}, N_∞, and τ, for the YA and OA populations.

In Section 3, we reprocess our data to remove the exponential global trend within each individual's CFT thus uniformizing different time scales at which different individuals' retrieval processes operate. This nonlinear procedure permits us to see the local structure of the inter-call times independently of the individuals' exponential exhaustion rates. Only then we can justify pooling the de-trended inter-call times to study their intrinsic statistical structure within the data sets of statistically more significant sizes. The exploratory nonparametric inference is conducted in Section 4, and the parametric analysis involving the Weibull distribution is conducted in Section 5.

Finally, Section 6, discusses our conclusions, their relevance to the rich literature of the subject, and the work in progress. The original data, and the *Mathematica* program codes used in the paper, are available, by request, from the corresponding author.

1.4 Methods

Subjects Subjects were young adults, enrolled as undergraduates at Case Western Reserve University, and older, cognitively normal adults participating in a longitudinal research registry at the Oregon Health Sciences University in Portland, OR, or the Memory and Cognition Center at University Hospitals Case Medical Center in Cleveland, OH. All subjects provided signed consent.

Testing procedures All subjects were tested in accordance with the procedures described previously. For this study, non-animals and repeated items were not excluded from analysis but were less than 1% of responses. The animal naming task was recorded and each audio file was scored for time of onset of each word using Apple GarageBand software (Apple, Inc., Cupertino, CA). Inter-rater reliability was 99%. The time between onset of words is called here the *inter-call (iC) time* and the average of the two raters was used in subsequent analyses described herein. The Case data were collected by the three first-named authors, at that time enrolled as undergraduate students at Case, and all the data were rated by them.

2 Experimental data and their exponential global structure

We have conducted animal Category Fluency Tests for 17 healthy young adults (YA, all volunteer students at Case Western Reserve University), and 17 healthy older adults (OA, ages 45–65)). For each individual we have recorded the exact timing of their N_{60} responses and estimated the model parameters in Eq. (1.4): N_∞, representing individual's "total recall capacity" allowing infinite time for recall, and τ, which is the exponential time "latency" constant. The latter can be conveniently thought of as the time by which the individual reaches $e^{-1} = 36.8\%$ of his "total recall capacity".[1]

Remark We must emphasize again that the parameter N_∞ is called here the *"total recall capacity"* only figuratively, with quotation marks applied advisedly. The actual recall process cannot possibly extend its exponential behavior to infinite time as a matter of both mathematics and common sense. Accepting the unlimited exponential behavior would practically mean that after, say, one hour the individual's recall ability would be essentially zero, an obvious nonsense. So N_∞ is just a useful parameter in the Rohrer model.

Our estimation procedure for the above two parameters does not use Rohrer's 10-s-bins histograms. For them, the constant N_∞ represents the asymptotic value at $t = 0$ which, given the relatively small data size for each individual, is notoriously hard to estimate in our context. Also, the binning procedure leads to some

[1] It would be more elegant to work here with the exponential function of the form $2^{-t/\tau}$ (instead of $e^{-t/\tau}$) in which case the "latency constant", τ, would be the time by which the individual reaches $2^{-1} = 50\%$ of his total capacity, but we elected not to do it as such a choice would make other calculations more messy.

loss of information contained in the original data set. Instead, we directly deal with the cumulative recall function $R(t)$, and our estimation makes an essential use of all the data points, $0 = t_0 < t_1 < t_2 < t_3 < \cdots < t_N$. It is based on the exponentially penalized least-squares method which we devised just for this purpose. It is described in the following mathematical aside.

2.1 Penalized least-squares fit of the cumulative count function

Our goal is to find parameters N_∞, and τ, guaranteeing the best least-squares fit between the sequence

$$R(t_k) = N_\infty(1 - e^{-t_k/\tau}), \qquad k = 1, 2, \ldots, N_{60},$$

and the sequence, $k = 1, 2, \ldots, N_{60}$. To simplify our calculations we will operate with parameter $\nu = 1/\tau$ rather than τ itself. The first instinct is to minimize the plain quadratic deviation function

$$L_0(N_\infty, \nu) = \sum_{k=1}^{N_{60}} (R(t_k) - k)^2 = \sum_{k=1}^{N_{60}} \Big(N_\infty(1 - e^{-t_k\nu}) - k\Big)^2.$$

For a perfect fit, shown in Fig. 1 below, the above approximation error would be zero. This straightforward method puts equal weights on the early recall times and the late ones. However, due to the exponential exhaustion, there are many more of the former so that the information contained in the latter is underutilized in the estimation process, a bad thing for judging an exponential decay. For the ideal example of exact exponential growth shown in Fig. 1 this does not matter. But having used the plain least squares method for real CFT data often resulted in a nonsensical estimate of N_∞ that was smaller than the number N_{60} of responses in the actual CFT 60-s test.[2] To remedy this difficulty we have increased weights of later responses exponentially and thus the goal became to find parameters N_∞, and ν, which minimize the *penalized quadratic deviation function*,[3]

$$L(N_\infty, \nu) = \sum_{k=1}^{N_{60}} e^{2t_k\nu}(R(t_k) - k)^2 = \sum_{k=1}^{N_{60}} \Big(N_\infty(e^{t_k\nu} - 1) - ke^{t_k\nu}\Big)^2. \tag{2.1}$$

Solution of the above minimization problem requires finding N_∞, and ν, satisfying the following two *normal equations*,

$$\frac{\partial L}{\partial N_\infty} = 2\sum_{k=1}^{N_{60}} \Big(N_\infty(e^{t_k\nu} - 1) - ke^{t_k\nu}\Big) \cdot (e^{t_k\nu} - 1) = 0, \tag{2.2}$$

and

$$\frac{\partial L}{\partial \nu} = 2\sum_{k=1}^{N_{60}} \Big(N_\infty(e^{t_k\nu} - 1) - ke^{t_k\nu}\Big) \cdot (N_\infty - k)t_k e^{t_k\nu} = 0. \tag{2.3}$$

Solving the first equation, Eq. (2.2), gives the estimator $\hat{N}_\infty$ as a function of ν,

$$\hat{N}_\infty(\nu) = \frac{\sum_{k=1}^{N_{60}} k \cdot e^{t_k\nu}(e^{t_k\nu} - 1)}{\sum_{k=1}^{N_{60}} (e^{t_k\nu} - 1)^2}. \tag{2.4}$$

After substitution of $\hat{N}_\infty(\nu)$ into Eq. (2.3) the estimator $\hat{\nu}$ for ν can be found by solving the equation,

$$\Lambda(\nu) := \sum_{k=1}^{N_{60}} \Big(\hat{N}_\infty(\nu)(e^{t_k\nu} - 1) - ke^{t_k\nu}\Big) \cdot (\hat{N}_\infty(\nu) - k)t_k e^{t_k\nu} = 0 \tag{2.5}$$

Clearly, there is no obvious analytic solution for Eq. (2.5), but one can find the first approximation to its root graphically. In Fig. 2, we are showing the plot of the function $\Lambda(\nu)$ in the neighborhood of its root in the

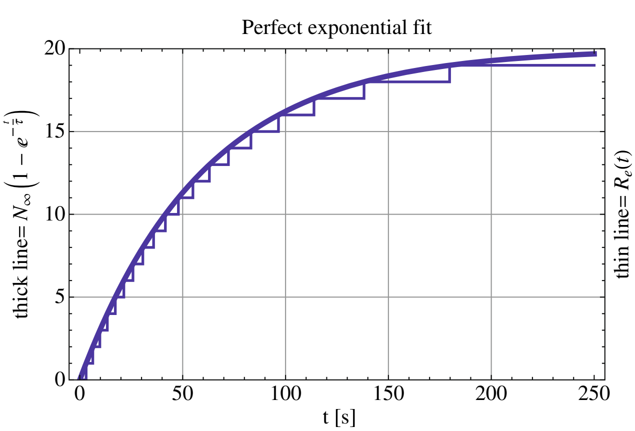

Fig. 1 An example of a perfect fit between the cumulative recall process $R_e(t) = k$, for $t_k \le t < t_{k=1}$, (cf. Eq. (1.4)), and the exponential function $N_\infty(1 - \exp(-t/\tau)$. Here, $N_\infty = 20$, and $\tau = 60$, and N, at $t = 60$, equals approximately 12

[2] Even worse, as a result, the crucial de-trending process discussed in Section 3 sometimes resulted in negative inter-arrival times.

[3] The coefficient 2 in the discounting exponential factor has been chosen for mathematical convenience so that we can distribute the two factors of $e^{t_k\nu}$ into the squared term in parentheses.

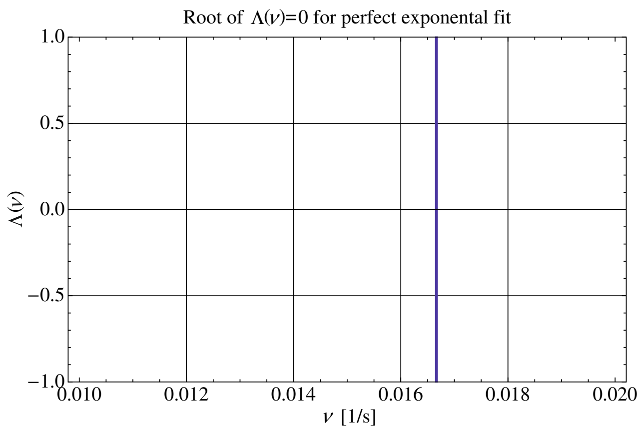

Fig. 2 A plot of the function $\Lambda(\nu)$ from Eq. (2.5). The root $\hat{\nu}$ of the equation $\Lambda(\nu) = 0$ is clearly about 0.017. *Mathematica*'s `FindRoot` algorithm then gives $\hat{\nu} = 0.0167$ which, in turn, yields an estimator for τ, $\hat{\tau} = 1/\hat{\nu} = 59.88$

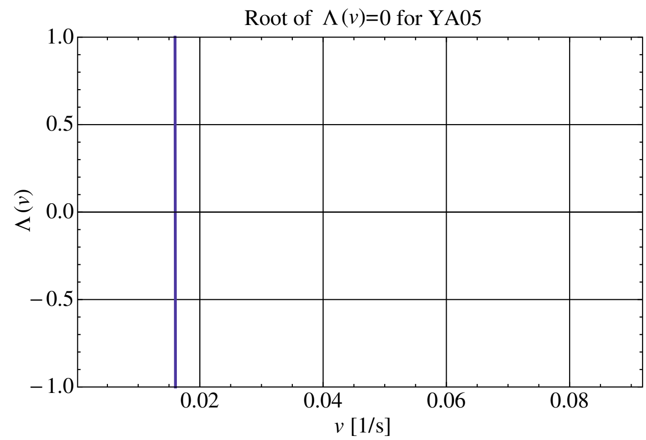

Fig. 3 A plot of the function $\Lambda(\nu)$ from Eq. (2.5) in the case of data from the younger adult YA05. The root $\hat{\nu}$ of the equation $\Lambda(\nu) = 0$ is just above 0.015. *Mathematica* `FindRoot` algorithm then gives $\hat{\nu} = 0.0148$ which gives an estimator for τ, $\hat{\tau} = 1/\hat{\nu} = 67.4$

ideal case shown in Fig. 1.[4] That first approximation can then be used as a seed for the *Mathematica* numerical algorithm `FindRoot` to find the actual value of $\hat{\nu}$.

Inspecting Fig. 2, we immediately find that the root $\hat{\nu}$ of the equation $\Lambda(\nu) = 0$ is about 0.017. *Mathematica* `FindRoot` algorithm then gives $\hat{\nu} = 0.0167$ which, in turn, yields an estimator for τ, $\hat{\tau} = 1/\hat{\nu} = 59.88 \approx 60$. Substituting this value into formula (2.4) returns $\hat{N}_\infty = 19.97 \approx 20$, both excellent estimates of the original parameters used in this simulation. In the next subsection we will move from the above "toy example" of our estimation procedure in the ideal world of perfect exponentials to the analysis of real data.

2.2 Estimating "total recall capacities" and the time "latency" constants for YA and OA populations

Having described and verified our estimation procedure for the parameters, N_∞, and τ, in the exponential model (1.3), we are now ready to apply it to the data collected from 17 young adults (labeled, YA01, ..., YA17), and 17 older adults (labeled, OA01, ..., OA17).

For each individual, the CFT was administered asking each individual to name, within a 60-s time period, as many animals as they could recall. Each session was recorded on Apple's MacBook Pro laptop using the *Garage Band* software which turned out to be perfectly suited for the purpose as it provided accurate timing of each sound recorded.[5] The timing of the beginning of each recalled name was read off the recording manually providing us with the recall sequence

$$0 < t_1 < t_2 < t_3 < \cdots < t_N < 60, \qquad (2.6)$$

specific for each individual. The beginning of the test was marked as $t_0 = 0$. The rough original data are available from the corresponding author, WAW, by request.

To illustrate how the algorithm from Section 2.1 works for the experimental data, we have selected the older adult YA05, for whom the total number of responses in the CFT was $N_{60} = 32$, and the sequence of recall times (1.1) was

```
1.4, 2.3, 2.8, 4, 4.8, 5.8, 7.8, 8.6,
9.7, 10.2, 11.4, 12.5, 13.9, 21.7,
23.4, 24.5, 25.6, 27.5, 28.4, 31.9,
32.3, 33.9, 37.8, 39.8, 41.8, 43.7,
47.6, 48.9, 52.8, 54.6, 55.6, 56.5
```

For this set of recall times the plot of the function $\Lambda(\nu)$ is shown in Fig. 3. Here the root $\hat{\nu}$ of the equation $\Lambda(\nu) = 0$ is clearly about 0.015. The formal *Mathematica* `FindRoot` algorithm then gives $\hat{\nu} = 0.0148$ which yields an estimator for τ, $\hat{\tau} = 1/\hat{\nu} = 67.4$. Substituting this estimate into Eq. (2.4) we obtain $\hat{N}_\infty = 54.6$.

The fit between the cumulative recall process $R_e(t) = k$, for $t_k \le t < t_{k=1}$, (cf. Eq. (1.4)) for YA05, and

[4] All computing and graphing in this paper has been done in the symbolic manipulation language *Mathematica*.

[5] The idea of using the *Garage Band* originated with our student coauthors (JM, and TS) who administered actual CFTs for young adults YA 01–17, and older adults OA 01–17. The data for OA 01–09 were supplied by another of our coauthors (JK).

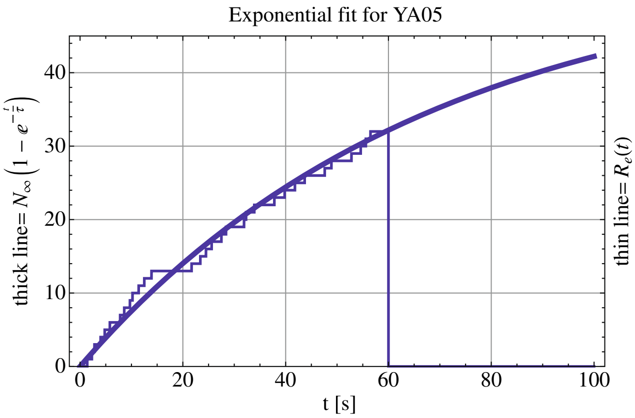

Fig. 4 The fit between the cumulative recall process $R_e(t) = k$, for $t_k \le t < t_{k=1}$, (cf. Eq. (1.4)) for YA05, and the model exponential function (1.3) $R(t) = N_\infty(1 - \exp(-t/\tau))$, with $N_\infty = 51.8$, and $\tau = 109$

the model exponential function (1.3), $R(t) = N_\infty(1 - \exp(-t/\tau)$, with $N_\infty = 54.6$, and $\tau = 67.4$, obtained by the above estimation algorithm, is shown in Fig. 4. The plot of the empirical $R_e(t)$ ends at $t = 60$ s, but, for the sake of the illustration, we extrapolated the theoretical model $R(t)$ up to $t = 100$ by which time $R(t)$ reaches about 58% of the "total recall capacity", N_∞. Here, for the first time, we will note the "burstiness" of the recall sequence data. It is a "second-order" effect related, presumably, to "category switching" which will be formally analyzed under our model in subsequent sections. For an in-depth, semantics based study of the problem, see Troyer et al. (1997), and also our concluding remarks in Section 6.

The inter-call sequences for all tested YA and OA individuals were then run through the above estimation procedure. The resulting estimates for the parameters, N_∞, and τ, are presented in Table 1 together with the original word counts, N_{60}.

The table shows that all three global parameters discriminate well between the YA and OA populations. The Student's T-test p-values for the difference between YA and OA, being zero, for $\hat{N}_{60}$, $\hat{N}_\infty$, and $\hat{\tau}$, are respectively $<$ .00007, .0014, and .012. All three global parameters are significantly higher for the YA population than for the OA population.

Another interesting observation is, that the coefficients of variations σ/μ, for parameters $\hat{N}_\infty$, and $\hat{\tau}$, for the YA are much higher than those for the OA, whereas, for N_{60}, the same coefficient remains relatively low (at around 0.2), and similar, for both YA, and OA. Simply put, the YA population has higher word counts, higher "total recall capacity", and higher "latency" constants (i.e., shows less "exponential exhaustion").

The compressed visual representation of the data in Table 1 is shown in Fig. 5 in the form of the so-called Box-and-Whiskers Plots for the estimators N_{60}, $\hat{N}_\infty$, $\hat{\tau}$ of the three global parameters. Recall that the standard Box-and-Whiskers Plot creates a graphical display with a box that spans the distance between two quartiles surrounding the median with the vertical lines

Table 1 Estimates of the "total recall capacities" N_∞, and time "latency" constants τ, for young adults (YA) and older adults (OA) in tested subpopulations

YA	N_{60}	$\hat{N}_\infty$	$\hat{\tau}$	OA	N_{60}	$\hat{N}_\infty$	$\hat{\tau}$
YA01	21	51.8	109	OA01	14	25.0	69.3
YA02	29	43.6	56.1	OA02	21	27.0	34.5
YA03	21	112	227	OA03	20	25.3	39.1
YA04	23	37.4	66.4	OA04	14	20.7	27.6
YA05	32	54.6	67.4	OA05	11	14.0	33.4
YA06	26	28.5	23.8	OA06	24	34.0	52.2
YA07	36	110	153	OA07	23	28.8	38.9
YA08	21	36.9	69.5	OA08	12	13.6	29.4
YA09	24	45.7	70.8	OA09	18	24.6	40.1
YA10	32	67.4	92.9	OA10	26	45.1	65.9
YA11	23	28.2	29.6	OA11	20	25.0	38.3
YA12	20	21.3	20.9	OA12	15	20.1	28.4
YA13	17	21.7	35.9	OA13	15	16.2	22.1
YA14	29	34.9	33.8	OA14	13	14.1	23.5
YA15	22	30.2	42.1	OA15	15	19.5	41.7
YA16	27	54.0	67.3	OA16	17	22.0	37.6
YA17	26	31.3	35.8	OA17	21	36.2	70.6
μ_{YA}	25.2	47.6	70.7	μ_{OA}	17.6	24.2	40.7
σ_{YA}	5.1	26.9	52.7	σ_{OA}	4.6	8.47	15.2
σ_{YA}/μ_{YA}	0.20	0.56	0.75	σ_{OA}/μ_{OA}	0.25	0.35	0.37
$\min_{YA}$	17	21.3	20.9	$\min_{OA}$	11	13.6	22.1
$\max_{YA}$	36	112	227	$\max_{OA}$	26	45.1	70.6

The original word counts N_{60}, as well as the means, standard deviations, coefficients of variations, minimum and maximum statistics for YA and OA subpopulations are also included in the table

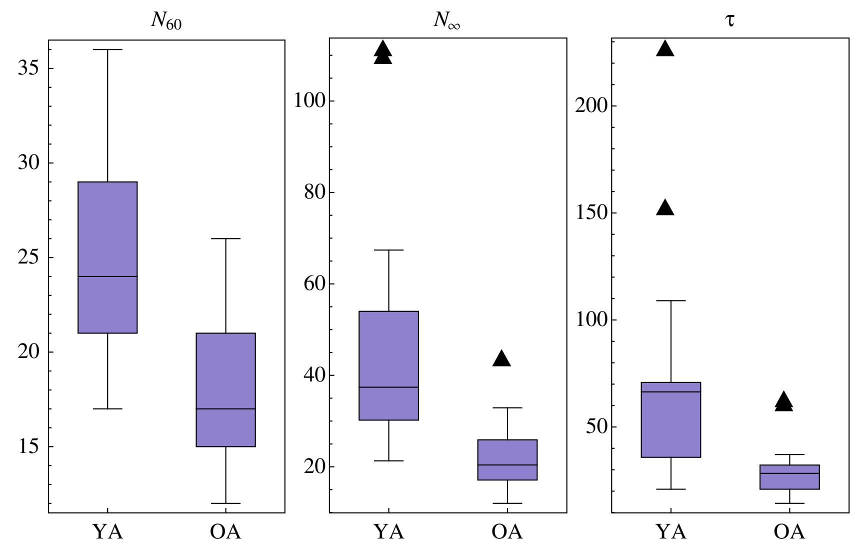

Fig. 5 Each of the three graphics compares the Box-and-Whiskers Plots for the YA (on the *left*) and the OA (on the *right*) populations: the left-most graphic, for the word count N_{60}, the *center*, for the parameter $\hat{N}_\infty$, and the *right*, for the parameter $\hat{\tau}$. For each of the three global parameters, the interquartile boxes for YA and OA populations do not overlap, reinforcing our conclusion that the differences between the global parameters for those two populations are significant

("whiskers") that extend to span the data set excluding outliers. Then near outliers (marked here by triangles) are defined as points beyond 3/2 times the interquantile range from the edge of the box. Far, or extreme, outliers (marked by squares) are defined as points beyond three times that range.

Each of the three graphics in Fig. 5 compares the Box-and-Whiskers Plots for the YA (on the left), and the OA (on the right) populations: the left-most graphic, for the word count N_{60}, the center, for the parameter $\hat{N}_\infty$, and the right, for the parameter $\hat{\tau}$. For each of the three global parameters, the interquartile boxes for YA and OA populations do not significantly overlap, reinforcing our conclusion that the differences between the global parameters for those two populations are significant.

3 Detrending the recall sequence

To study the intrinsic statistical properties of the random inter-call times $\delta t_1, \ldots, \delta t_N$, $N = N_{60}$, we need to remove the individual's exponential exhaustion (corresponding to different time scales at which each individual's retrieval processes operate) from the random recall times $t_1, \ldots, t_N$ (see Eq. (1.2)). Otherwise, the inter-call times, say, $\delta t_1, \ldots, \delta t_{12}$, in Fig. 1, cannot be thought of as representing random quantities with similar probability distribution, and be used in the derivation of the common underlying statistical characteristics of the inter-call times, which we feel represent the essential information about the retrieval-from-semantic-memory process.

The detrended recall sequence, $T_1, \ldots, T_N$, is designed to grow approximately linearly as if the responses were produced at a steady pace throughout the test at the rate of one response per unit of the (new, rescaled) time. Then the detrended random inter-call times, $\delta T_1, \ldots, \delta T_N$, can be reasonably thought of (again, within the area of validity of the Rohrer's

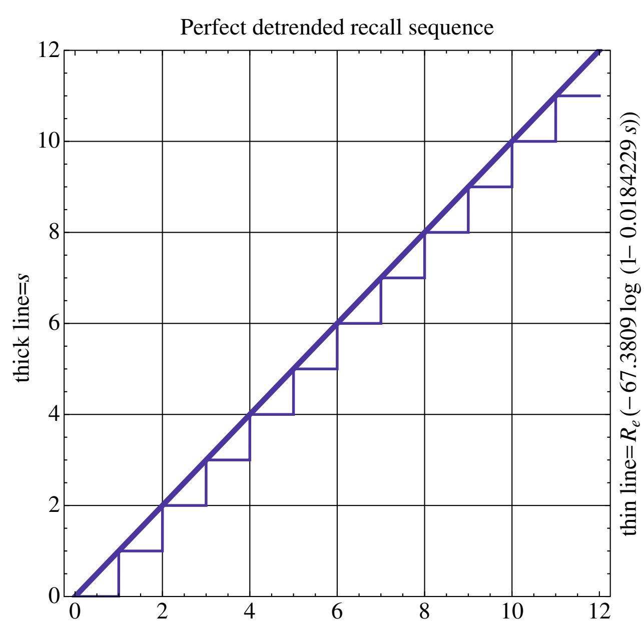

Fig. 6 The ideal exponential case. The cumulative response function in the original, natural time t, was shown in Fig. 1. Here, the same empirical cumulative response count $R_e(t)$, is shown as a function $R_e(t(s))$ of the new, nonlinearly rescaled time s. It grows linearly at the rate of one response per unit of the (new) time. Here, and in Fig. 1, $N_{60} = 12$

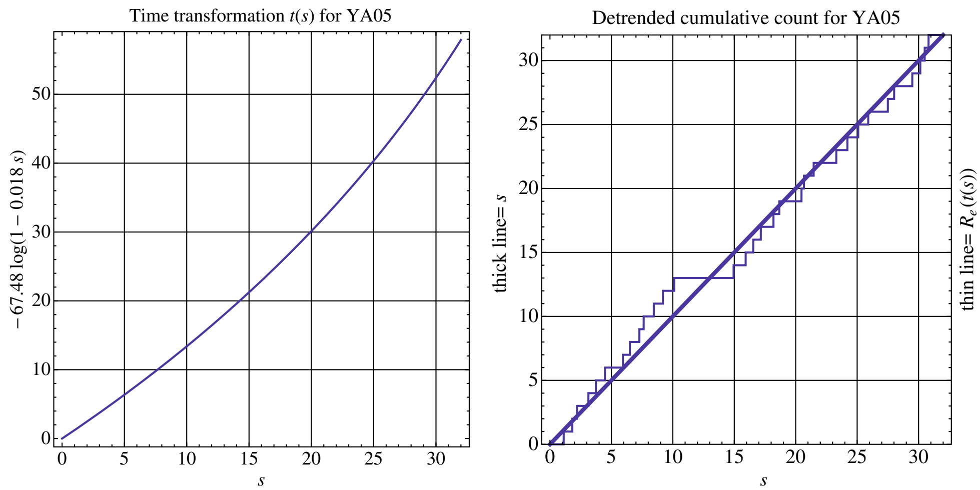

Fig. 7 *Left* The time transformation, $t = t(s)$, for data YA05 shown in Fig. 4. In this case, $N_{60} = 32$. *Right* The resulting empirical cumulative response count $R_e(t(s))$ for YA05, as a function of the new, nonlinearly rescaled time s

exponential model) as being similarly distributed. In the ideal nonrandom case shown in Fig. 1 their size would be exactly equal to 1. The latter ensemble of approximately identically distributed, and independent[6] random quantities can then be used for statistical inference about their distributional properties. The above properties will also justify aggregation of the detrended data across the YA and OA populations to increase sample sizes in our statistical inference procedures.

The algorithm implementing the detrending process is as follows: The goal is to find the nonlinear time-stretching function $t = t(s)$ such that $t(0) = 0$, and $t(N_{60}) = 60$, so that the detrended exponential response function $R(t(s))$ becomes a linear function of the new time s. This leads to the equation

$$N_{60} \frac{1 - e^{-t(s)/\tau}}{1 - e^{-60/\tau}} = s,$$

which can be easily solved for $t(s)$, giving us the logarithmic time transformation,

$$t(s) = -\tau \log\left(1 - \frac{s}{N_{60}}(1 - e^{-60/\tau})\right). \tag{3.1}$$

The normalization condition, $t(N_{60}) = 60$, standardizes the rescaled time so that it increases by one unit with each response in the idealized exponential case shown in Figs. 1 and 6.[7]

For data YA05 shown in Fig. 4, the time transformation $t = t(s)$ is shown in Fig. 7(left). The resulting detrended empirical cumulative response count $R_e(t(s))$, as a function of the new, nonlinearly rescaled time s, is shown in Fig. 7. In this case, $N_{60} = 32$. Again, even after the detrending process, the cumulative response count function shows burstiness which we will discuss later on.

At this point we will apply the above detrending algorithm to the times-of-recall sequences, $t_1, \ldots, t_{60}$, for each of the tested individuals, YA1 through YA17, and OA1 through OA17, thus producing random detrended inter-call times sequences $\delta T_1, \ldots, \delta T_{N_{60}}$, for each of them. In the next section we will provide a preliminary nonparametric comparison of their statistical properties.

4 Exploratory nonparametric analysis of detrended inter-call (dTiC) times

Comparison of probability distributions of detrended intercall times for different individuals at a respectable significance level is limited by the small sizes, N_{60}, of the samples, typically around 20. Here is where our detrending processing comes in handy, justifying aggregation of all YA, and OA, δT data. This creates aggregated samples of detrended intercall (dTiC) times of size $N_{YA} = 412$, for the young adults, and $N_{OA} = 291$,

[6] Some sort of independence, perhaps fairly week, within the sequence $\delta t_1, \ldots, \delta t_N$, and thus, $\delta T_1, \ldots, \delta T_N$, is needed to justify employment of the standard statistical inference tools here. This is a delicate issue as consecutive responses are probably *not* completely statistically independent, and the problem is related to the *global temporal structure* of the category fluency tests. On the other hand, one can reasonably assume that the detrended inter-call sequences for different individuals *are* statistically independent.

[7] Note that the information about the total number of responses in each test is already embedded in the usual parameter N_{60}, we do not need it anymore.

for the older adults. For ideal data showing perfect "exponential exhaustion" the means of the YA, and OA, δT samples should both be standardized to 1. For our data the mean for the YA data turns out to be 0.963, and that for the OA data, 0.971. Both of them are reasonably close to 1 given the empirical nature of our data obtained from widely varying human subjects.

So, at this point of our study, our goal is to compare more subtle distributional properties for dTiC times for the YA, and OA populations, and the obvious first step is a construction of the Q-Q plot for them. The latter is shown in Fig. 8. Recall, that it is the 2-D parametric plot of the curve,

$$(Q_{YA}(p), Q_{OA}(p)), \qquad 0 < p < 1,$$

where $Q_{YA}(p)$, and $Q_{OA}(p)$, are, respectively, the quantile functions for the aggregated YA, and OA data.

The first surprising observation is that that the YA and OA data have similar distributions for small values of dTiC times. The cutoff point is about 1.3, and about 77% of the data fall below it. Above that point the distributions diverge. This leads us to the hypothesis that the small values, presumably corresponding to dTiC times between animal names within the same category, are produced by a different mechanism than the large values corresponding to intercall times related to switching between different animal categories. This definition, resulting from the purely statistical exploration of the dTiC data, makes us independent of the particular definition of what constitutes a semantic category. The remainder of the paper will discuss the above formal hypothesis using various statistical evidence. For the sake of convenience, formally, but admittedly somewhat arbitrarily, we shall refer to the times less than 1.3 as *intra-category dTiC times* and those larger than 1.3 as *inter-category dTiC times*. The worthwhile investigation of the question to what extent our formal classification corresponds to the actual semantic category switching as defined and studied, for example, by Troyer et al. (1997), Pollio et al. (1969) and Graesser and Mandler (1978), is in our future plans; also, see Section 6.

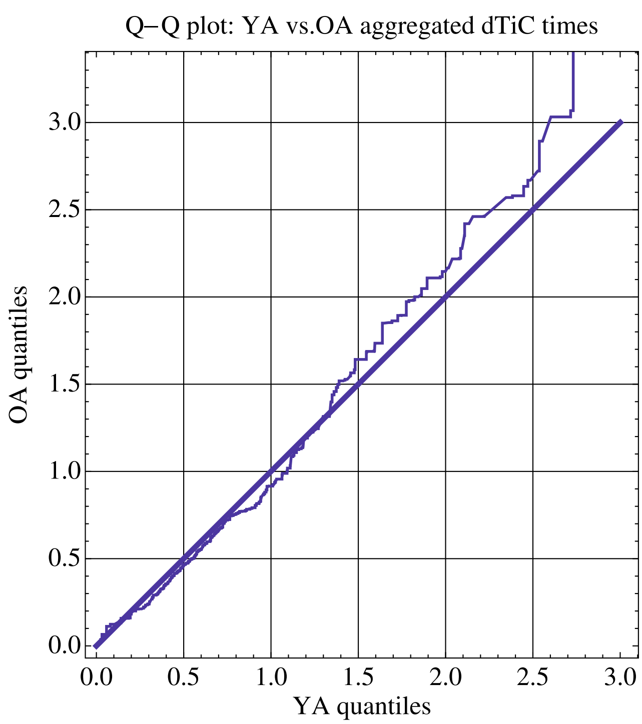

Fig. 8 The Q–Q-plot comparison of the OA and YA aggregated dTiC times data. For values up to approximately 1.3 the two distributions are similar whereas they diverge for larger values. The phenomenon seems to be related to the animal category switching. This suggests that the dTiC data smaller than the cut-off point 1.3 are produced by a different recall mechanism than those above it. If the two data sets were identically distributed, the Q–Q plot would follow the thick diagonal straight line with slope 1

In Fig. 9 we show plots of the empirical cumulative distribution functions (CDFs) of intra-category dTiC times (left top) and inter-category dTiC times (right top) for YA (thin lines), and OA (thick lines). The deviations between CDFs in both cases between YA and OA populations are pictured directly beneath the corresponding CDF plots. The maximum absolute deviation for the inter-category CDFs is about two times larger than that for the intra-category CDFs. Moreover, the intra-category OA CDF is located to the left of that of YA, while the opposite is true for the inter-category CDFs.

More formally, denoting by

$$F_{YA<1.3}(t), \quad F_{OA<1.3}(t), \quad F_{YA>1.3}(t), \quad \text{and} \quad F_{OA>1.3}(t),$$

the empirical CDFs of, respectively, aggregated intra-category dTiC times for YA,, and OA, and inter-category dTiC times for YA, and OA (shown in Fig. 9), and by

$$N_{YA<1.3}, \quad N_{OA<1.3}, \quad N_{YA>1.3}, \quad \text{and} \quad N_{OA>1.3},$$

the corresponding populations' sizes, the Kolmogorov–Smirnov statistics (see, e.g., Denker and Woyczynski 1998) for the hypothesis of equality of two empirical distributions are

$$K_{<1.3} = \sqrt{\frac{N_{YA<1.3} \cdot N_{OA<1.3}}{N_{YA<1.3} + N_{OA<1.3}}} \cdot \max_t \left| F_{YA<1.3}(t) - F_{OA<1.3}(t) \right| \approx 1.03$$

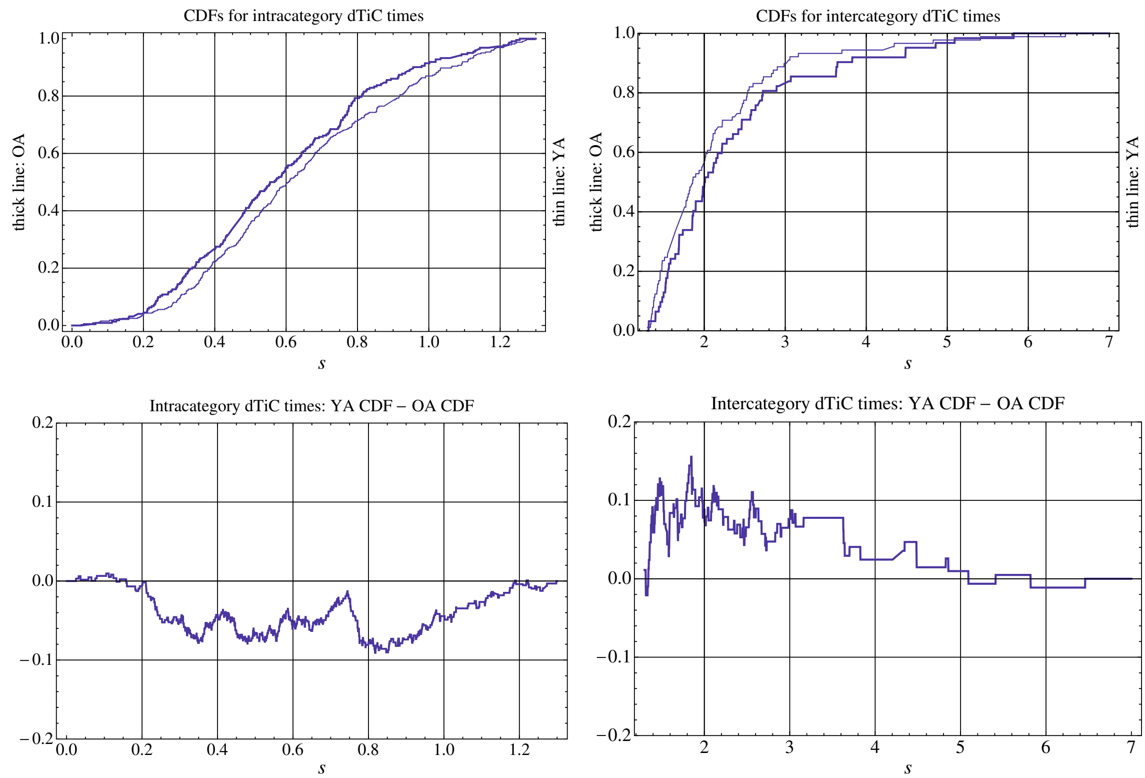

Fig. 9 Plots of the empirical cumulative distribution functions (CDFs) of aggregated intra-category dTiC times (*left top*) and inter-category dTiC times (*right top*) for YA (*thin lines*), and OA (*thick lines*). The deviations in both cases between YA and OA populations are pictured directly beneath the corresponding CDF plots. The maximum absolute deviation for the inter-category CDFs is about two times larger than that for the intra-category CDFs. Moreover, the intra-category OA CDF is located to the left of that of YA, while the opposite is true for the inter-category CDFs

for the intra-category dTiC times, and

$$K_{>1.3} = \sqrt{\frac{N_{YA>1.3} \cdot N_{OA>1.3}}{N_{YA>1.3} + N_{OA>1.3}}} \cdot \max_t \left| F_{YA>1.3}(t) - F_{OA>1.3}(t) \right| \approx 1.15$$

for the inter-category dTiC times. At 15% significance level the critical value for the Kolmogorov–Smirnov test is $K_{cr}(0.15) = 1.14$, so that we have

$$K_{<1.3} < K_{cr}(0.15) < K_{>1.3}.$$

Hence the conclusion is that the hypothesis of the equality of the distributions of the inter-category dTiC times for YA and OA populations,

$$H_0 : F_{YA>1.3}(t) = F_{OA>1.3}(t),$$

can be rejected, while a similar hypothesis for intra-category dTiC times for YA and OA populations,

$$H_0 : F_{YA<1.3}(t) = F_{OA<1.3}(t),$$

cannot be rejected.

However, at 10% significance level the critical value for the Kolmogorov–Smirnov test is $K_{cr}(0.10) = 1.22$, so that we have

$$K_{<1.3} < K_{>1.3} < K_{cr}(0.10).$$

Hence at that significance level the conclusion is that the hypothesis of the equality of the distributions of the inter-category dTiC times for YA and OA populations cannot be rejected for either the inter-category or inter-category dTiC times, On the other hand, at 25% significance level, when the critical value for the Kolmogorov–Smirnov test is $K_{cr}(0.25) = 1.02$, the equality hypotheses can be rejected for both, inter-catogory and intra-category cases. So, the situation is delicate and we will study this problem in the next section via more quantitative parametric tools.

The more subtle effects of dominations of CDFs observed in Fig. 9 (top) correspond to the formal probability theory concept of strong domination of random quantities, see Kwapien and Woyczynski (1992, p. 111). Recall, that if X and Y are two positive random quantities, then X is said to be strongly dominated by Y, if for

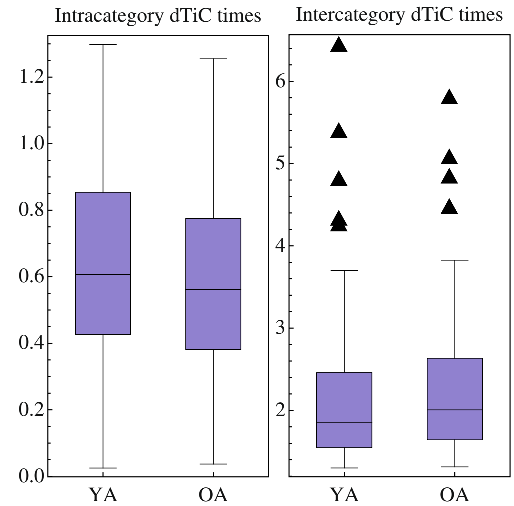

Fig. 10 The Box-and-Whiskers plot comparisons of the YA (plot 1) and OA (plot 2) aggregated dTiC times data. Again, the evidence is strong that the intra-category dTiC times for YA subpopulation dominate those for the OA subpopulations (plot in the *left figure*). The reverse is true for the inter-category dTiC times (plot in the *right figure*). Also, note the presence of numerous outliers for the inter-category dTiC times, and their absence for intra-category dTiC times

each $t \geq 0$, the tail probabilities of Y are greater than the tail probabilities of X, that is,

$$\mathbf{Pr}(X > t) \leq \mathbf{Pr}(Y > t),$$

or, equivalently, $F_X(t) \geq F_Y(t)$ for the CDFs, and $Q_X(p) \leq Q_Y(p)$, for the corresponding quantile functions. Thus, in view of Fig. 9, considered as random quantities, the *YA intra-category dTiC times strongly dominate the OA intra-category dTiC times*, and the reverse is true for the inter-category dTiC times. However, the effect is slight for the intra-category times and quite pronounced in the inter-category data.

The Box-and-Whiskers plots comparisons of the OA and YA aggregated dTiC times data in Fig. 10 reinforce the above conclusions.

5 A parametric model for detrended inter-call times: Weibull probability distribution

5.1 Why Weibull distribution?

Our hypothesis is that the detrended inter-call times, $\delta T_1, \ldots, \delta T_{N_{60}}$, follow the Weibull probability distribution. We have arrived at this hypothesis by some exploratory data analysis but also being guided by the likely nature of the process of recall from the semantic memory. Here, our admittedly simplistic, "first-past-the-post" model[8] corresponds to the picture of multiple names (or categories) stored in the semantic memory starting, at the beginning of each recall cycle, a race through the neural network of the brain. The one with best (smallest, always random) "arrival time" wins by being called out.[9]

The above picture of the name recall process corresponds to the following formal statistical model: Consider a sequence of independent random variables, $\Delta_1, \ldots, \Delta_N$, with positive values identically distributed with a CDF, $F_\Delta(t)$, and define a new random quantity

$$\Delta_{\min,N} = \min_{1\leq i\leq N} \Delta_i. \tag{5.1}$$

Note that we have returned here to denoting the time variable by the letter t, although from now onwards it is the nondimensional detrended time.

In view of the independence assumption, the tail cumulative distribution function

$$\begin{aligned} 1 - F_{\Delta_{\min,N}}(t) &= \mathbf{Pr}(\Delta_{\min,N} \geq t) = \mathbf{Pr}(\min_{1\leq i\leq N} \Delta_i, \geq t) \\ &= \mathbf{Pr}^N(\Delta_1 \geq t) = (1 - F_\Delta(t))^N, \end{aligned} \tag{5.2}$$

which depends both on N, and the initial distribution, $F_\Delta(t)$, of each $\Delta_1, \ldots, \Delta_N$. However, it is a remarkable result in Large Sample Theory (for a simple proof, see, e.g., the compactly written book by Ferguson 1996) that, under a mild restriction demanding a power decay of the tails[10] of the original c.d.f. $F_\Delta(t)$, there exists a *universal* (that is, independent of $F_\Delta(t)$) limit distribution of the minima, as the sample size $N \to \infty$. It is of the form

$$\begin{aligned} W(t) &= F_{\Delta_{\min,\infty}}(t) = \lim_{N\to\infty} F_{\Delta_{\min,N}}(t) \\ &= 1 - e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}, \quad \text{for} \quad t > \gamma > 0, \end{aligned} \tag{5.3}$$

and equal to zero for $t \leq \gamma$, with the corresponding p.d.f.

$$\begin{aligned} w(t) &= f_{\Delta_{\min,\infty}}(t) \\ &= \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}, \quad \text{for} \quad t > \gamma > 0, \end{aligned} \tag{5.4}$$

[8] A more subtle parametric model that automatically takes into account the burstiness of the recall process will be studied in another paper.

[9] We have found recently, that a similar argument was used by Logan (1995) in the context of what he calls the "instance theory of automaticity".

[10] This property is often called the heavy-tail property, and it means that the probabilities of big values are much larger than the corresponding probabilities for Gaussian or exponential distributions.

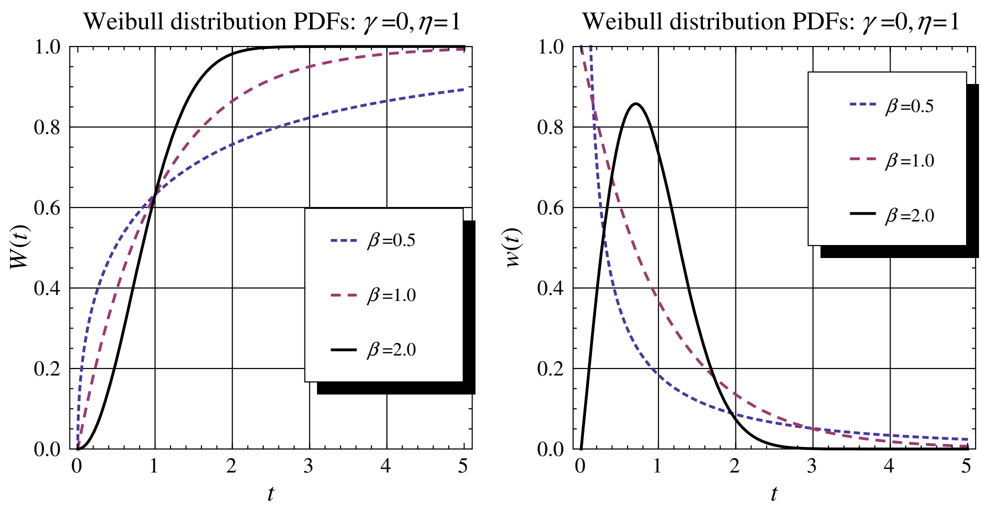

Fig. 11 Examples of CDFs (*left*) and the corresponding PDFs (*right*) of the Weibull distribution for the following selected values of the shape parameter: $\beta = 0.5$, 1, and 2, (*thin, dashed*, and *thick lines*, respectively). Parameter γ is set at zero, and $\eta = 1$

and zero for $t \le \gamma$, and it is known as the Weibull distribution. Parameter β is usually called the *shape* parameter, and η is called the *scaling* parameter. Given its analytic shape, the Weibull distribution is sometimes called the *stretched exponential distribution*; for $\beta = 1$ it becomes just the standard exponential distribution.

Weibull distribution first appeared in studies of reliability of complex devices, see Weibull (1951). It is not hard to see why. Think of a ball bearing with N balls, each subject to independent random wear and thus having an independent random lifetime before failing. The whole ball bearing fails when the first ball in the device wears out. Thus the lifetime of the whole ball bearing is the minimum of the lifetimes of independent balls it incorporates. There is a vast body of work applying Weibull distribution in different contexts, the most relevant for us being studies of the random access times in searches from computerized data bases (see, e.g., Shirani-Mehr et al. 2008).

The Weibull distribution has three parameters. Parameter $\gamma > 0$ is the cutoff point; the probability of values below γ is zero. For us it is an important quantity as it represents the minimal time needed for recall in each individual (or group of individuals). We will call it *reaction time.* Parameter η permits a change of scale. One would have to adjust it when one wants to change time measurement units from, say, seconds to minutes. With detrended standardized inter-call data, it will play a lesser role in what follows; intuitively speaking it should be not too far to 1, given that the mean value of the Weibull distribution is $\eta\Gamma(1 + 1/\beta)$, with parameter β varying in our data between 1 and 2. Finally, β is the key parameter affecting the shape of the Weibull distribution and the qualitative behavior of the corresponding recall process, especially as β crosses the $\beta = 1$ threshold. It should be observed that any Weibull random quantity can be obtained from an exponential random quantity via the stretching transformation $t \mapsto t^\beta$, a useful observation for simulations of the Weibull data. Figure 11 shows examples of CDFs, and PDFs of the Weibull distribution for selected values of the shape parameter β.

5.2 Parametric estimators for Weibull distribution: general principles

There exists a vast literature (mostly generated by the reliability theory research in engineering) on parametric inference for the Weibull distribution. We have tried several approaches and settled on the maximum likelihood estimation (MLE) procedure which was pioneered by Leone et al. (1960);[11] for more recent work, see. e.g., Thoman et al. (1969) and Wu (2002).

For a sample of size N of dTiC times, $\delta T_1, \ldots, \delta T_N$, the estimator $\hat{\gamma}$ for the cut-off parameter γ is

$$\hat{\gamma} = \min_{1 \le i \le N} \delta T_i. \tag{5.5}$$

This value will then be subtracted from each data point so that in what follows, we will simply assume that $\gamma = 0$, and the estimation problem is then reduced to the two parameters, β and η. In this case, see, e.g., Wu (2002), the maximum likelihood estimator $\hat{\beta}$ for the shape parameter β is a solution of the following transcendental equation,

$$\frac{1}{\hat{\beta}} + \frac{1}{N}\sum_{i=1}^{N} \log \delta T_i - \frac{\sum_{i=1}^{N} \delta T_i^{\hat{\beta}} \log \delta T_i}{\sum_{i=1}^{N} \delta T_i^{\hat{\beta}}} = 0, \tag{5.6}$$

[11] Coincidentaly, the paper was written fifty years ago when Fred Leone was director of the Statistical Laboratory here at Case.

Table 2 Estimates of the Weibull distribution parameters of intra- and inter-category dTiC times, for for young adults (YA), and older adults (OA), in tested subpopulations

YA	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\eta}$	OA	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\eta}$
$YA < 1.3$	0.02	2.41	0.72	$OA < 1.3$	0.04	2.33	.067
$YA > 1.3$	1.30	1.01	0.84	$OA > 1.3$	1.31	1.08	1.06

which, unfortunately, can only be solved numerically. Then the MLE $\hat{\eta}$ for the scaling parameter η can be calculated from the formula

$$\hat{\eta} = \left(\frac{1}{N} \sum_{i=1}^{N} \delta T_i^{\hat{\beta}} \right)^{1/\hat{\beta}} . \tag{5.7}$$

5.3 Estimates of Weibull parameters for the empirical dTiC times

In view of our preliminary nonparametric exploration of the dTiC in Section 4, the above estimation process has been then implemented[12] for the four aggregated data sets:

(a) the aggregated YA, and OA, intra-category dTiC times, denoted, respectively $YA < 1.3$, and $OA < 1.3$, and
(b) the aggregated YA, and OA, inter-category dTiC times, denoted, respectively $YA > 1.3$, and $OA > 1.3$.

The results are shown in Table 2, below.

The parametric estimates in Table 2 reflect the domination effects mentioned in Section 4. For intra-category times,

$$\hat{\beta}_{YA<1.3} = 2.41 > \hat{\beta}_{OA<1.3} = 2.33,$$

$$\hat{\eta}_{YA<1.3} = 0.72 > \hat{\eta}_{OA<1.3} = 0.67.$$

YA parameters dominate the OA parameters but the differences are slight. The reverse inequalities are true for the inter-category times,

$$\hat{\beta}_{YA>1.3} = 1.01 < \hat{\beta}_{OA<1.3} = 1.08,$$

$$\hat{\eta}_{YA<1.3} = 0.84 < \hat{\eta}_{OA<1.3} = 1.06.$$

and the differences are more pronounced.

The differences are most striking between the intra- and inter-category estimates for the crucial shape parameter β indicating, perhaps, different mechanisms of the recall from the semantic memory in those two cases. Indeed, $\hat{\beta}$'s in the intra-category case hover around 2.4 whereas they are close to 1 for the inter-category data. The latter fact gives rise to an interesting observation: For inter-category data for both, YA and OA populations, the distributions of dTiC times are close to those in the most elementary Poisson process-like exponential model. To a lesser extent the results also indicate a more subtle difference between how category switching operates in younger adults and older adults.

In Fig. 12 we show parametric Weibull CDF fits (thick lines) for the the empirical CDFs of aggregated YA intra-category dTiC times (left top), OA intra-category dTiC times (right top), YA inter-category dTiC times (left bottom), and OA inter-category dTiC times (right bottom). The inter-category CDFs have been shifted to the left by 1.3 to allow better comparison of their shapes with those of the intra-category CDFs. The empirical CDFs are plotted as thin lines. The plots show good fits between the theoretical Weibull distributions and the empirical data. The dramatic differences between the intra-category ($YA < 1.3$, $OA < 1.3$) and inter-category ($YA > 1.3$, $OA > 1.3$) CDFs reflect the parametric estimates shown in Table 2.

Finally, to conclude this section, we provide in Table 3 MLEs for the Weibull parameters γ, β, and η, for the dTiC times for all the individuals, YA 1–17, and OA 1–17. This was done without separating the intra-category and inter-category dTiC times for each individual; the sample sizes seemed too small for a meaningful statistical analysis. Although the data in Table 2 are interesting in their own right, in view of the small sample size, the reliability of these estimators is somewhat limited; note the large values of the calculated coefficients of variation σ/μ. But the calculation of the mean values of the parameters still shows that the shape parameter β for the YA population dominates, on the average, that for the OA population,

$$\mu_{YA}(\hat{\beta}) = 1.29 > \mu_{OA}(\hat{\beta}) = 1.07$$

thus reinforcing its value in the study of local properties of the CFT process.

Figure 13 shows the Box-and-Whisker Plots for the local parameters from Table 2 for the tested individuals in the YA (the left plot in each of the three graphics) and the OA (the right plot) populations. The left-most graphic shows the picture for the reaction time parameter $\hat{\gamma}$, the center, for the Weibull shape parameter $\hat{\beta}$, and the right, for the Weibull scaling parameter $\hat{\eta}$. In contrast to the Box-and-Whiskers plots for the

[12] Here we used the `FindRoot` facility in *Mathematica*.

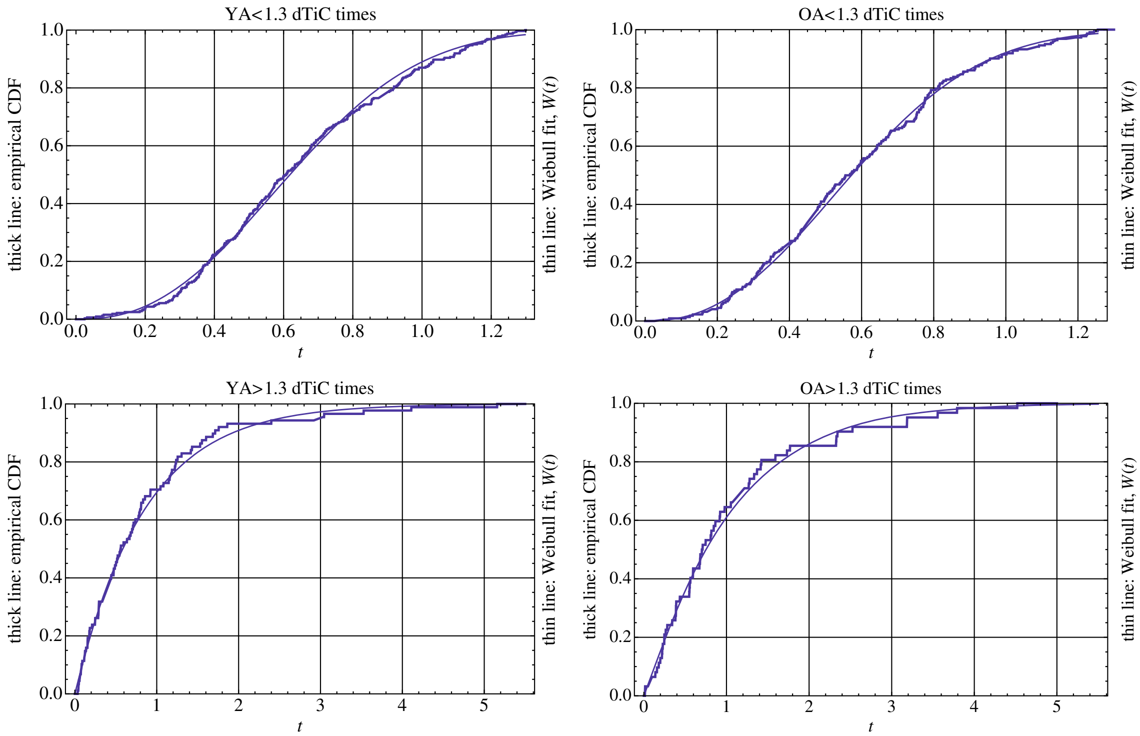

Fig. 12 Parametric Weibull CDF fits (*thick lines*) for the the empirical CDFs of aggregated YA intra-category dTiC times (*left top*), OA intra-category dTiC times (*right top*), YA inter-category dTiC times (*left bottom*), and OA inter-category dTiC times (*right bottom*). The empirical CDFs are plotted as *thin lines*. The inter-category CDFs have been shifted to the left by 1.3 to allow better comparison of their shapes with those of the intra-category CDFs. The plots show good fits between the theoretical Weibull distributions and the empirical data. The dramatic differences between the intra-category and inter-category CDFs reflect the parametric estimates shown in Table 2

Table 3 Estimates of the Weibull distribution parameters of detrended inter-call ($dTiC$) times sequences, for individual young adults (YA), and older adults (OA), in tested subpopulations

YA	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\eta}$	OA	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\eta}$
YA01	0.11	1.22	1.00	OA01	0.11	0.75	0.79
YA02	0.02	1.13	1.03	OA02	0.23	1.01	0.76
YA03	0.14	1.04	0.77	OA03	0.44	0.90	0.58
YA04	0.02	1.11	1.05	OA04	0.33	1.04	0.49
YA05	0.20	1.20	0.84	OA05	0.20	0.69	0.67
YA06	0.07	1.87	1.06	OA06	0.24	0.91	0.78
YA07	0.29	1.33	0.80	OA07	0.12	1.21	0.99
YA08	0.33	1.25	0.75	OA08	0.07	0.99	1.05
YA09	0.12	1.18	0.91	OA09	0.24	0.59	0.42
YA10	0.04	1.16	1.01	OA10	0.16	2.02	0.98
YA11	0.25	1.43	0.80	OA11	0.09	0.99	0.98
YA12	0.11	2.08	1.04	OA12	0.50	1.07	0.41
YA13	0.36	1.25	0.75	OA13	0.15	1.73	1.07
YA14	0.21	1.28	0.90	OA14	0.24	0.88	0.84
YA15	0.29	1.31	0.88	OA15	0.12	1.36	1.00
YA16	0.21	1.11	0.71	OA16	0.21	1.12	0.86
YA17	0.14	1.00	0.91	OA17	0.29	1.05	0.77
μ_{YA}	0.17	1.29	0.89	μ_{OA}	0.22	1.07	0.79
σ_{YA}	0.11	0.28	0.12	σ_{OA}	0.12	0.36	0.21
σ_{YA}/μ_{YA}	0.65	0.21	0.13	σ_{OA}/μ_{OA}	0.53	0.26	0.27
$\min_{YA}$	0.02	1.00	0.71	$\min_{OA}$	0.04	0.59	0.41
$\max_{YA}$	0.36	2.08	1.06	$\max_{OA}$	0.50	2.02	1.07

Included at the bottom are their summary statistics: means, standard deviations, coefficients of variation, minima, and maxima

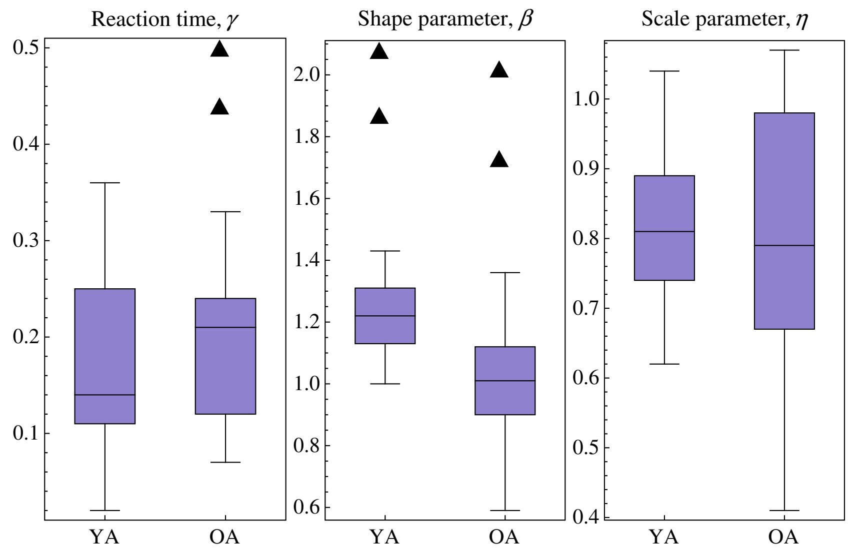

Fig. 13 The Box-and-Whiskers Plots for the local parameters from Table 2 for the tested individuals in the YA (the *left plot* in each of the three graphics) and the OA (the *right plot*) populations. The *left-most* graphic shows the picture for the reaction time parameter $\hat{\gamma}$, the *center*, for the Weibull shape parameter $\hat{\beta}$, and the *right*, for the Weibull scaling parameter $\hat{\eta}$. In contrast to the Box-and-Whiskers plots for the global parameters shown in Fig. 5, the interquartile boxes for the estimators of local parameters $\hat{\gamma}$, and $\hat{\eta}$ largely overlap, and it is only the shape parameter $\hat{\beta}$ that clearly distinguishes between the YA and OA populations

global parameters shown in Fig. 5, the interquartile boxes for the estimators of local parameters $\hat{\gamma}$, and $\hat{\eta}$ largely overlap, and it is only the shape parameter $\hat{\beta}$ that clearly distinguishes between the YA and OA populations.

6 Concluding remarks, future work

Our study of the local random structure of the One Minute Animal Naming Task permits a more detailed analysis of the differences between various statistical properties of CFT response processes for younger and older adult populations as well as the nature of the differences between probability distributions for what we called (for lack of better words) intra-category and inter-category responses. Clearly, subtly different mechanisms are involved in the retrieval from semantic memory in the intra-category and inter-category situations. We worked within the "exponential exhaustion" model developed by Rohrer et al. (1995). It worked reasonably well for our data and permitted the detrending operation which gave us a larger, statistically more significant pool of "homogenized" aggregated data to work with.

It is important to re-emphasize that our goal *was not* to find the best possible analytic fit for our data as this could be trivially accomplished by using expressions with more and more parameters; it would have no scientific significance. Instead we sought a relatively simple model providing some rationale for the mechanism of the recall processes. Here, the Weibull model introduced in Section 5 seemed natural to us in the context of CFT and relied on the common-sense intuition that recall from semantic memory involves a supply of "names" which are searched independently of each other and the one with the minimum search time is the one retrieved. The three-parameter Weibull distribution helped us quantify the above mentioned differences. We settled here on the maximum likelihood estimation of Weibull parameters but other methods are possible, and we also experimented with the linear regression fit in the log-log scales; it gave similar results to the MLE method.

Roughly speaking the basic conclusions are:

(i) The probability distributions of the intra-category inter-call times dominate those of the inter-category times;
(ii) The probability distributions of the younger adult inter-call times dominate those of the older adult times; and
(iii) The Weibull distribution has a simple justification and is a reasonably well-fitting choice describing statistical properties of inter-call times, while the Weibull shape parameter β is a sensitive parameter which could be used to evaluate the differences mentioned in (i) and (ii).

However, one should remember that the testing protocols were only as rigorous and consistent as was possible given a variety of human participants. So, the (three-digit) accuracy of the numbers appearing in our tables is somewhat illusory and should be taken with a grain of salt. What matters are the differences between various phenomena exhibited by those numbers.

Finally, we would like to place our work in the context of the rich literature on the subject of probability distributions of inter-call times. First, the issue of terminology. In mathematical and experimental psychology papers, various terms have been employed to describe similar concepts; reaction times in e.g. McGill and Gibbon (1965), inter-response times (IRT) in, e.g., Rouder et al. (2008), Van Zandt (2000), or latency times in Ratcliff and Murdock (1976). We settled on a different, communication theory influenced term "inter-call (iC) times" because our precise definition signified time intervals between the start of the pronunciation of consecutive words (onsets) rather than the time intervals between the end of the pronunciation of the previous word (offset) and the onset of the next word, as used for example by Murdock and Okada (1970). Also, for anybody studying the recall problems, an extensive review of the empirical literature in field by Wixted and Rohrer (1994) should be required reading.

The Weibull model is, of course, not the only one that has some justification here although some of the more recent work by Cousineau et al. (2002) strengthens the reasoning behind its use as an approximate distribution of the minimum even in the case when the individual random quantities are not identically distributed. Also Rouder et al. (2008) analyzed a hierarchical approach for fitting curves to the response times measurements using Weibull distribution although their motivation was different from ours.

From the very beginning of our work we have also considered "building a better mousetrap" and model the recall-from-semantic-memory-process as a random walk process with a positive drift. Our main modeling effort developed in this paper was based on a simple idea: the item recalled first is the one that wins in the competition for the shortest recall times between all the items stored in the semantic memory, sort of, first-past-the-post concept. A more subtle model can be based on the following idea: once the effort to recall an item from the semantic memory is initiated different competing items perform a "random walk", or "diffusion", in the "subconscious" domain (or neural network) reinforced by a positive drift of our effort to get the item into the "conscious" domain where it could finally be pronounced. The item that crosses first a certain threshold between the "subconscious" and "conscious" domains is the one named first.

Mathematically, in the simplest form, the above reasoning calls, in the continuum limit, for the analytic description by the stochastic diffusive process,

$$X(t) = a \cdot t + \sigma \cdot B(t), \qquad t > 0, \tag{6.1}$$

where $a > 0$ is a deterministic drift coefficient, $\sigma > 0$ is the diffusion coefficient, and $B(t)$ is the continuous-time random walk normalized by the condition Var $B(1) = 1$ (the standard Brownian motion process). It is well known (see, e.g., Sashadri 1993) that, for a given threshold level, $L > 0$, the probability distribution function $f_\theta((t)$ of the (random) time $\theta = \min\{t > 0 : X(t) = L\}$, when the process $X(t)$ crosses the level L for the first time, is of the form

$$f_\theta(t) = \frac{L}{\sqrt{2\pi\sigma^2}} t^{-3/2} \exp\left(-\frac{(at-L)^2}{2\sigma^2 t}\right), \qquad t > 0.$$

The above distribution is known as the *Inverse Gaussian*, or *Wald*, PDF. The mean m_θ, and the variance σ_θ^2 of the first level-L crossing time θ are given by the formulas,

$$\mu_\theta = \frac{L}{a}, \qquad \sigma_\theta^2 = \frac{L\sigma^2}{a}. \tag{6.2}$$

In the next step of our crude experimentation with the random walk model we calculated the empirical mean and variance of our aggregated young adult detrended intercall time intervals obtaining the mean value 0.078 and variance 0.48. Substituting these values into Eq. (5.2) we obtained the following estimates:

$$\hat{\sigma}^2 = .62, \qquad \hat{a} = 1.28,$$

for parameters in the Inverse Gaussian distribution (the parameter L was set, arbitrarily, at $L = 1$).[13] We did not separate the inter-category and intra-category data. The resulting fit, as well as the 90 percent confidence bands are shown below in Fig. 14 (left), and the corresponding Inverse Gaussian p.d.f., in Fig. 14 (right).

The Inverse Gaussian fit of Fig. 14 shows the accuracy not dissimilar to the one we obtained previously for the Weibull distribution. In a related experiment, see Fig. 15, below, we made a direct comparison of the deviations of the empirical CDF of the detrended iC times, $\delta T_1, \ldots, \delta T_{N_{60}}$ (aggregated for all Young Adults (YA)) from Weibull CDF fit (left), and the corresponding Inverse Gaussian CDF fit (right). The

[13] In statistics, this approach to parameteric estimation is called the Method of Moments.

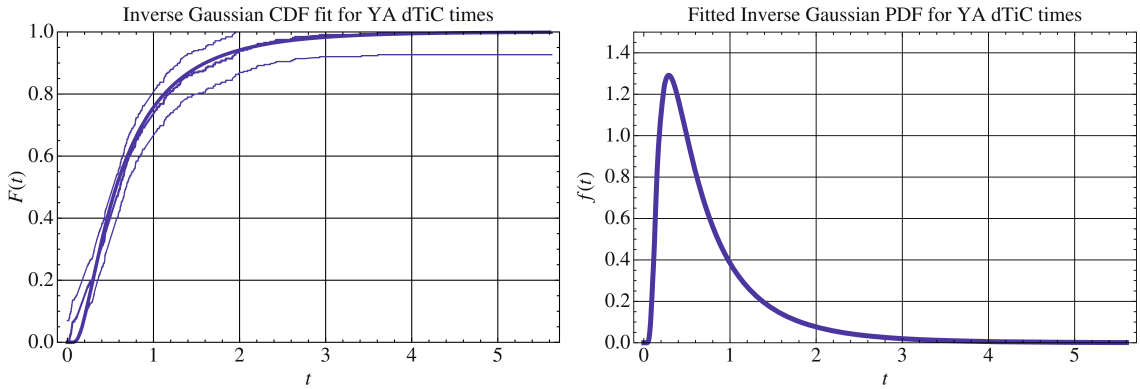

Fig. 14 *Left* The 90-% confidence band for the Kolmogorov–Smirnov goodness-of-fit test for the Inverse Gaussian CDF fitted to dTiC times $\delta T_1, \ldots, \delta T_{N_{60}}$ aggregated for all Young Adults (YA). *Right* The PDF of the fitted Inverse Gaussian CDF shown on the *left*

accuracies in both cases are similar although the Weibull fit performed better for small values of the detrended inter-call time intervals, and the Inverse Gaussian fit was slightly better for the large ones. Could that be a sign that intra-category recall process is governed by the "first-past-the post", Weibull model, and the inter-category switching by the random walk, Inverse Gaussian model?

As it turned out, thanks to the information provided by the referees who reviewed the first version of this paper, we were just rediscovering the magisterial work on the diffusion-based theory of memory retrieval developed in 1978 by Ratcliff (1978), and pursued subsequently by him and his collaborators, see, e.g., Ratcliff and Murdock (1976), Ratcliff et al. (1999, 2004), and most recently White et al. (2010).

In another direction, Rohrer (1996) provided a good fit of a portion of the IRT distribution by using a mixture of a Gaussian and an exponential distribution; and Gamma distribution makes its appearance in the work of McGill and Gibbon (1965) as a natural distribution of the cumulative response times once the exponential distribution is assumed for the individual IRTs. Also, in 2007, Rhodes and Turvey (2007) suggested modeling human memory retrieval as Lévy foraging. The use of Lévy, and, in particular Lévy α-stable processes, $0 < \alpha < 2$, for modeling anomalous jump-diffusive random dynamics, has a long, and honorable tradition in science and engineering, see. e.g., a review by Woyczynski (2001), and the literature cited therein, and the above mentioned authors argue for its appropriateness in the retrieval-from-memory context. Without expressing an opinion about the physiological merits of the model we would like to point out that the α-stable cumulative distributions, $F_\alpha(x)$, have the probability tails decaying at the power rate $1/x^\alpha$, so that the corresponding minima, see Section 5.1, are also asymptotically approximated by the Weibull distributions. Thus there is some commonality in the two fits. We refrained from running those models for our data because, for reasons explained above, the race for the best fit was not on our agenda.

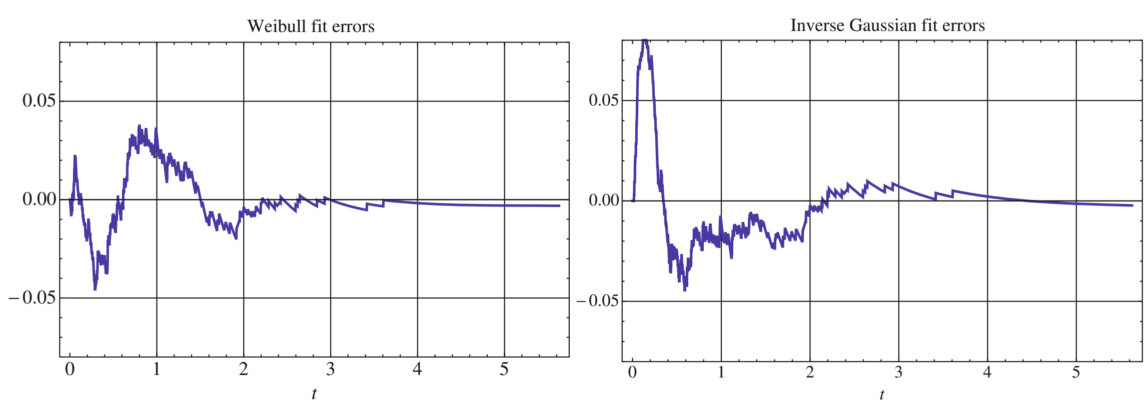

Fig. 15 Deviations of the empiirical CDF of the detrended iC times, $\delta T_1, \ldots, \delta T_{N_{60}}$ (aggregated for all Young Adults (YA)) from the fitted Weibull CDF (*left*), and from the fitted Inverse gaussian CDF (*right*). The accuracies in both cases were similar

There are several interesting issues that we would like to pursue in the future.

(a) Our designation of “inter-category”, and “intra-category” iC times was purely formal and mathematical. So the natural next step is to investigate to what extent these “inter-category” switching times correspond to the actual semantic category switching as studied, for example, by Troyer et al. (1997), Pollio et al. (1969) and Graesser and Mandler (1978). We are currently doing some preliminary work in this direction but the problem does not seem to have a unique and well defined solution as any definition of what constitutes a category within the individual's semantic memory is contestable. Our initial efforts gave somewhat scattered results depending on what types of categories were considered (say, habitat, vs. taxonomy). However, the problem is outside the scope of this paper.

(b) Also, there is a question of the “burstiness” of the process which may, or may not be related to category switching. Our classification of intra-category and inter-category times is only the first and somewhat superficial step. But it is possible to construct other seamless models that would incorporate those effects. The Weibull model with shape parameter $\beta < 1$ does display a bursty, intermittent behavior but our estimates of the shape parameters were greater than 1. A very interesting analysis of the burstiness vs. category switching is contained in a 2002 paper on the dynamics of memory retrieval in older adulthood by Wingfield and Kahana (2002). We see here an opening for stochastic models based on the more general Markov processes that were employed in the past in traffic studies where the bunching-up behavior is the essential part of the dynamics of the phenomenon.

(c) A more futuristic exploration would combine the Brownian motion-based random walk approach originated by Ratcliff with the Lévy nonlocal jump diffusion used by Rhodes and Turvey while including some nonlinear effects. We believe there is a good physiological argument for such an approach; shock waves that can appear in such models can explain some of the burstiness in the data and the category-switching paradigm. The nontrivial mathematics for such an approach has been developed only recently but is available, see, e.g., Woyczynski (2005), Piryatinska et al. (2005), Karch and Woyczynski (2007) and Jourdain et al. (2008). An even more ambitious project, would consider the diffusions of multidimensional (or even infinite-dimensional) manifolds. Physiologically, given the structure of the neural network in the brain, this approach may be even more justified than all the previously considered models. The mathematics here has been developed for a while, see, e.g. Ellworthy (1982), but is rather intricate. Applying it to and interpreting it for real data will be a daunting task.

Out hope is that the above sensitive techniques can be now used to evaluate individuals with various form of dementia. We did not have a large enough collection of data from, for example, Alzheimer patients to carry this out but also plan to do this in the future.

In brief, compared to previous efforts, our work includes a systematic use of nonparametric approach, employs the tool of detrending to create larger pools of inter-call data in the context of Category Fluency Tests, and uses Weibull parametric model justified by a very simple recall from semantic memory model.

Acknowledgements This research has been supported by the U.S. National Science Foundation Grant “Interdisciplinary Training for Undergraduates in Biological and Mathematical Sciences” (DUE-0634612) administered by PJT; the first three authors were undergraduate students in biology and mathematics at Case while this work has been done. PJT was supported by the National Science Foundation's program in Mathematical Biology (DMS-0720142), and acknowledges research support from the Oberlin College Libraries. JK was supported by grants from the National Institute on Aging (P30AG024978, P30AG08017 and R01AG024059).

The authors are grateful to the anonymous referees for helping us understand the broad and rich history of the studies of the distributional properties of inter-response times about which which we were initially quite ignorant partly because of the terminological diversity in the relevant literature. The paper benefited from their kind and generous advice.

References

Bousfield, W. A., & Sedgwick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology, 30*, 149–165.

Canning, S. J., Leach, L., Stuss, D., et al. (2004). Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia. *Neurology, 24*, 556–562.

Caramelli, P., Carthery-Goulart, M. T., Porto, C. S., et al. (2007). Category fluency as a screening test for Alzheimer disease in illiterate and literate patients. *Alzheimer Disease and Associated Disorders, 21*, 6–67.

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

Denker, N., & Woyczynski, W. A. (1998). *Introductory statistics and random phenomena: Uncertainty, complexity and chaotic behavior in engineering and science*. Boston: Birkhauser.

Diaz, M., Sailor, K., Cheung, D., & Kuslansky, G. (2003). Category size effects in semantic and letter fluency in Alzheimer's patients. *Brain and Language, 89*, 108–114.

Ellworthy, K. D. (1982). *Stochastic differential equations on manifolds*. Cambridge: Cambridge University Press.

Ferguson, N. (1996). *Large sample theory*. London: Chapman and Hall.

Gomez, R. G., & White, D. A. (2006). Using verbal fluency to detect very mild dementia of the Alzheimer type. *Archives of Clinical Neuropsychology, 21*, 771–775.

Graesser, A., & Mandler, G. (1978). Limited processing capacity constrains the storage of unrelated sets of words and retrieval from natural categories. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 86–100.

Jourdain, B., Méléard, S., & Woyczynski, W. A. (2008). Nonlinear stochastic differential equations driven by Levy processes and related partial differential equations. *ALEA-Latin American Journal on Probability and Mathematical Statistics, 4*, 1–29.

Karch, G., & Woyczynski, W. A. (2007). Fractal Hamilton–Jacobi–KPZ equations. *Transactions of the American Mathematical Society, 360*(2007), 2423–2442.

Kramer, J. H., Nelson, A., Johnson, J. K., et al. (2006). Multiple cognitive deficits in amnestic mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders, 22*, 306–311.

Kwapien, S., & Woyczynski, W. A. (1992). *Random series and stochastic integrals: Single and multiple*. Boston: Birkhäuser.

Leone, F. C., Rutenberg, Y. H., & Topp, C. W. (1960). *Order statistics and estimators for the Weibull distribution. 1960 report no. 10.26*. Cleveland: Statistical Laboratory, Case Institute of Technology.

Lerner, A. J., Orgrocki, P. K., & Thomas, P. J. (2009). Network graph analysis of category fluency testing. *Cognitive Behavioral Neurology, 22*(1), 45–52.

Logan, G. D. (1995). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review, 102*(1.4), 751–756.

McGill, W. J., & Gibbon, J. (1965). The general-gamma distribution and reaction times. *Journal of Mathematical Psychology, 2*(1), 1–18.

Morris, J. C., Weintraub, S., Chui, H. C., et al. (2006). The Uniform Data Set (UDS): Clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Disorder and Associated Disorders, 20*, 210–216.

Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Experimental Psychology, 86*(2), 263–267.

Piryatinska, A., Saichev, A. I., & Woyczynski, W. A. (2005). Models of anomalous diffusion: The subdiffusive case. *Physica A: Statistical Mechanics and Applications, 349*, 375–420.

Pollio, H. W., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 529–536.

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

Ratcliff, R., McKoon, G., & Van Zandt, T. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.

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

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*(2004), 408–424.

Rhodes, T., & Turvey, M. T. (2007). Human memory retrieval as Lévy foraging. *Physica A, 385*, 255–260.

Rohrer, D. (1996). On the relative and absolute strength of a memory trace. *Memory and Cognition, 24*(2), 188–201.

Rohrer, D., Wixted, J. T., Salmon, D. P., & Butters, N. (1995). Retrieval from semantic memory and its implications for Alzheimer's disease. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2.1), 1127–1139.

Rouder, J. N., Tuerlinckx, F., Speckman, P., Lu, J., & Gomez, P. (2008). A hierarchical approach for fitting curves to response time measurements. *Psychonomic Bulletin and Review, 15*(6), 1201–1208.

Sashadri, V. (1993). *The inverse Gaussian distribution*. Oxford: Oxford University Press.

Sauzeon, H., Lestage, P., Raboutet, C., et al. (2004). Verbal fluency output in children aged 7–16 as a function of the production criterion: Qualitative analysis of clustering, switching processes, and semantic network exploitation. *Brain and Language, 89*, 192–202.

Shirani-Mehr, H., Li, C., Linag, G., & Shmueli-Scheuer, M. (2008). *Quality-aware retrieval of data objects from autonomous sources for web-based repositories*. http://www.ics.uci.edu/~chenli/pub/icde08-crawling.pdf.

Thoman, D. R., Bain, L. J., & Antle, C. E. (1969). Inferences on the parameters of the Weibull distribution. *Technometrics, 11*(3), 445–460.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14*, 167–177.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11*(1), 138–146.

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

Weibull, W. (1951). A statistical distribution of wide applicability. *Journal of Applied Mechanics, 18*, 293–297.

Weiner, M. F., Neubecker, K. E., Bret, M. E., & Hynan, L. S. (2008). Language in Alzheimer's disease. *Journal of Clinical Psychiatry, 69*(2.4), 1223–1227.

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology, 54*, 39–52.

Wingfield, A., & Kahana, M. J. (2002). The dynamics of memory retrieval in older adulthood. *Canadian Journal of Experimental Psychology, 56*(3), 187–199.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin and Review, 1*, 89–106.

Woyczynski, W. A. (2001). Lévy processes in the physical sciences. In T. Mikosch, O. Barndorff-Nielsen, & S. Resnick (Eds.), *Lévy processes—theory and applications* (pp. 241–266). Boston: Birkhauser.

Woyczynski, W. A. (2005). Nonlinear partial differential equations driven by Lévy diffusions and related statistical issues. In J. Duan, & E. C. Waymire (Eds.), *Probability and partial differential equations in modern applied mathematics, IMA* (Vol. 140, pp. 247–258). Berlin: Springer.

Wu, S.-J. (2002). Estimation of the parameters of the Weibull distribution with progressively censored data. *Journal of the Japan Statistical Society, 32*(2), 155–163.