Journal of Mathematical Psychology **46**, 431–454 (2002)
doi:10.1006/jmps.2001.1399

Extending Statistics of Extremes to Distributions Varying in Position and Scale and the Implications for Race Models

Denis Cousineau, Victor W. Goodman, and Richard M. Shiffrin

University of Indiana, Bloomington

Race models are characterized by the largest or smallest of samples from *n* distributions. The asymptotic theory of extremes has demonstrated that for identically distributed, independent, and lower-bounded random variables, whose left tail approximates a power function, the distribution of the minimum tends toward a Weibull distribution as *n* increases. In this article, we remove the restriction of identically distributed random variables by letting the lower bound or the scale of the random variables be random variables themselves. We prove that the Weibull distribution is still the asymptotic distribution of the minimum and relate its parameters to the parameters of the input distributions. We discuss the potential use of such findings in models of psychological processes.

INTRODUCTION AND DEFINITIONS

Race models assume a process in which a large number of units are in competition to determine the response. We will discuss races in which the fastest unit determines the response, although the results are mirrored in systems in which the slowest determines the response. Understanding race models therefore entails understanding the behavior of the minima of the samples.

Since the seminal works of Fisher and Tippett (1928), the distributions of extremes (minima and maxima) have been studied extensively. It has been established that if the minima as the number of competing units, *n*, goes to infinity have a distribution at all, the distribution will be of three possible types: Type I, the double-exponential distribution; Type II, the Fréchet distribution; or Type III,

Denis Cousineau is now at the Université de Montréal. This research was supported by a NIMH grant to Dr. Richard M. Shiffrin. We thank Gordon Logan and Christine Lefebvre for their comments on a previous draft of this article. Part of this research was presented at the 33rd Annual Meeting of the Society for Mathematical Psychology held in Kingston, OT, 2000.

Address correspondence and reprint requests to Denis Cousineau, Département de psychologie, Université de Montréal, C. P. 6128, succ. Centre-ville, Montréal, Québec H3C 3J7, Canada. E-mail: Denis.Cousineau@Umontreal.CA.

0022-2496/02 $35.00

the Weibull distribution. These attractors are commonly called the asymptotic distribution of extremes (Cramér, 1946; Feller, 1966; Leadbetter, Lindgren, & Rootzén, 1983). Type III can only occur when the units' distributions have a finite lower bound. Some authors have shown that race models producing Type-III distributions are related to Luce's choice model (Bundesen, 1990; Shibuya & Bundesen, 1988; Marley, 1989; Marley & Colonius, 1992).

In this article, we will focus on Type III for several subsidiary reasons: (i) the Weibull provides a good fit of response-time data (Logan, 1992; Cousineau, in preparation); (ii) the Weibull is closely related to the exponential family of distribution and, thus, to Poisson processes (Link, 1992; Wandell & Luce, 1978); (iii) the properties of the Weibull are rather well known (Weibull, 1951; Gumbel, 1958; Green & Luce, 1975). By far the most important reason, however, is our desire to use the present results for the modeling of psychological processes: Distributions of time for any component of a cognitive process must surely be restricted to positive values and hence must be lower bounded.

Before going on, we note that the minimum shrinks toward a single point as n approaches infinity. This is called degeneration. Thus the stable asymptotic results are concerned mostly with the shape of the asymptotic distribution, rather than with its scale or variance. As we shall see below, it is customary to normalize the sample of minima obtained from n units by a number depending on n, a number chosen so that the scale remains roughly constant as n increases. The issue of normalization has important implications for the modeling of psychological processes, but discussion of this is deferred to the third section.

In order to demonstrate the results of Sections 1 and 2, it is assumed that each unit is a random variable and that they are identically and independently distributed (i.i.d.). The i.i.d. assumptions are very restrictive, however, and are likely to be violated in many psychological race model theories.

For example, in the memory trace theory of Indow (1993) specific traces are retrieved through the use of cues. It is assumed that a trace becomes unretrievable in the experimental setting when any one of its supporting cues is lost. In this model, the loss of a trace corresponds to a race to determine when the first cue will fade. Indow assumed for the sake of simplicity that all links were identical and so the probabilities of losing a cue were all distributed equally. This assumption is unlikely, though, since some cues may be much stronger than others, an interpretation congruent with the primacy effect.

The instance-based model of Logan (1988) gives another example. Each encounter with a stimulus lays down a memory trace; retrieval with that stimulus as a cue produces a race among all past traces. Logan assumed equality among these traces, yet variability in storage seems more likely. For example, the first encounter with a stimulus may lay down a much more salient trace than the 100th encounter.

If the assumption of identical distributions is wrong, then predictions based on it could be wrong. It is therefore critical to learn in what ways the corresponding race models differ from those based on i.i.d. Thus this article explores the changes in the asymptotic distribution of extremes when the units racing are not identical.

In this paper, we will introduce variability across the units by manipulating two general parameters of the distribution functions, as seen in Fig. 1, namely position and scale. Although it is common to identify the position of a distribution by some measure of central tendency, such as the mean, it is more convenient in this article to define position as the location of the lower bound (defined as $\alpha(F)$ below). We shall later define the scale precisely; for now we may conceptualize the scale as something akin to standard deviation. The general idea will be to replace these values with random variables having certain properties, see if it is still possible to infer the asymptotic distribution of the minimum, and if so, verify whether this asymptotic distribution will be a Weibull.

Definitions and Conditions for I.I.D. to Converge toward a Weibull Distribution

In this section, we define most of the notions needed to understand the theorems presented next. We adopt some of the notations introduced in Galambos (1978).

Let $\mathbf{X}_i$ be a collection of i.i.d. random variables representing the performance for the n competing units and $\mathbf{X}_{1:n} := \min(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)$ be the response of the fastest competitor on a given trial. In the following equations, F denotes the distribution function of $\mathbf{X}_i$. Also, let the function α defined on a distribution function be

$$\alpha(F) := \inf\{x \mid F(x) > 0\}. \tag{0.1}$$

If a distribution function F is not lower-bounded, then $\alpha(F) = -\infty$. In order for $\mathbf{X}_{1:n}$ to converge toward a Weibull distribution, two criteria must be satisfied;

$$\begin{aligned} C_1(F) &:= \alpha(F) > -\infty \\ C_2(F) &:= \lim_{t \to \infty} \frac{F(\alpha(F) + 1/tx)}{F(\alpha(F) + 1/t)} = x^{-\gamma_F}, \end{aligned} \tag{0.2}$$

where $\gamma_F > 0$ and $x > 0$. It has been shown that C_1 and C_2 are necessary and sufficient for F to converge toward an asymptotic distribution of minima of Type III (e.g., Galambos, 1978).[1] The first criterion states that each unit's distribution must have a unique and finite lower bound $\alpha(F)$. The second criterion explores the form of the distribution when we, are in the left end of the distribution, that is, infinitely close to the lower bound $\alpha(F)$.

In essence, the criterion C_2 states that the left end of the distribution function F is *almost* a power curve with a positive exponent γ_F. The constant $\gamma_F > 0$ determines what will be the shape of the attractor distribution. The ratio form used in C_2 serves to bypass issues of scaling. A stronger criterion sometimes seen stipulates that F is *exactly* a power curve in the left tail,

$$C_{2a}(F) := \lim_{x \downarrow 0} \frac{F(\alpha(F) + x)}{x^{\gamma_F}} = K_F. \tag{0.3}$$

[1] Other forms for C_2 can be used. For example, Gnedenko (1943) proposed

$$C_{2'}(F) := \lim_{h \downarrow 0} \frac{F(hx - \alpha(F))}{F(h - \alpha(F))} = x^{\gamma}.$$

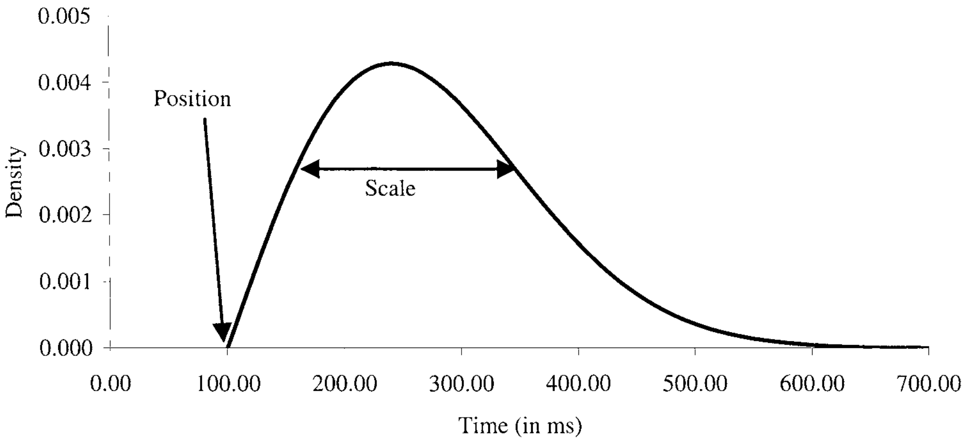

FIG. 1. Illustration of position and scale on a hypothetical lower-bounded distribution.

C_{2a} is more restrictive but is included in C_2, and all functions F satisfying C_{2a} also satisfy C_2 and thus have minima that are of Type III.[2] Many distributions used in psychology indeed satisfy C_{2a}, for example, the exponential, the Weibull, the Gamma, and the uniform distributions. Figure 2 shows examples of distributions of the first three of these families, along with an enlargement of their left-end tails. The dashed lines in the figure represent the power curve that fits the left-end tail, as seen in the enlargement on the bottom row. The power curve equation is obtained using Eq. (0.3) and is equal to $K_F x^{\gamma_F}$. It is easy to verify that these tails are power curves, and log–log plots will yield straight lines. The shape constant γ_F of the exponential and the uniform distributions is constant and equal to 1.

If both conditions C_1 and C_2 are met, then it is possible to formulate in closed form the attractor distribution. The asymptotic distribution of the minima, L, is defined by

$$L(F) := W(\gamma_F, \alpha(F), b_{F,n}),$$

where W is the Weibull distribution with position parameter $\alpha(F)$, scale parameter $b_{F,n}$, and shape parameter γ_F.[3] The cumulative density function (CDF) of the Weibull is given by

$$W(\gamma_F, \alpha(F), b_{F,n})\,(x) = 1 - e^{-\left(\frac{x-\alpha(F)}{b_{F,n}}\right)^{\gamma_F}}. \tag{0.4}$$

[2] An example of a distribution that does not satisfy C_{2a} but does satisfy C_2 is given by the function $F(x) = x^{\gamma} \log(e/x)$ defined over the domain $]0, 1[$. Clearly, the term x^{γ} satisfies a power curve, and the second term is shown to be a slowly-varying function (Feller, 1966) since

$$\lim_{t \downarrow 0} \frac{\log(e/tx)}{\log(e/t)} = 1.$$

Any product of a slowly-varying function with a true power function satisfies C_2. Another example of this type will be described in Section 3 (and Footnote 5).

[3] Galambos used the notation $L_{2,\gamma}$ to indicate that it is the second of three possible attractors. In other references, $L_{2,\gamma}$ is generally called the type-III distribution of extremes (e.g., Luce, 1986).

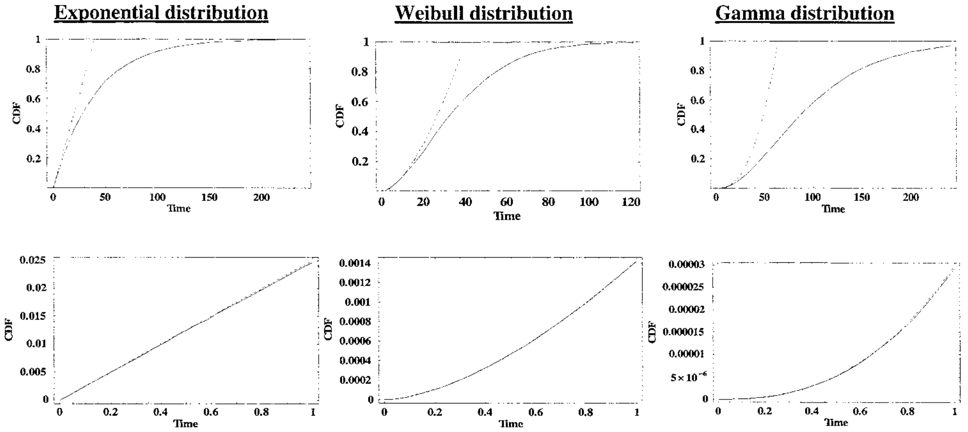

FIG. 2. Enlargement of the left-end tail of three distribution functions. All three functions have a lower bound $\alpha(F)$ equal to zero and arbitrary scales. In the top row are shown the CDF of an exponential distribution with scale parameter $b_E = 40$, a Weibull distribution with shape parameter $\gamma_W = 1.6$ and scale parameter $b_W = 40$, and a Gamma distribution with shape parameter $\gamma_G = 2.5$ and scale parameter $b_G = 40$. In the bottom row are shown enlargements of the left-end tails of the above distributions in the region [0, 1]. The dashed lines show the power curves that fit exactly the left-end tail of the three functions. The power curves are given by $K_F x^{\gamma_F}$, where $K_E = 1/b_E$, $K_W = (1/b_W)^{\gamma_W}$, and $K_G = 1/(\gamma_G! b_G{}^{\gamma_G})$. All Ks were obtained using Eq. (0.3).

As a special case, when $\gamma_F = 1$ the distribution becomes exponential, which recapitulates a well-known special case: minima sampled from exponential distributions are themselves exponential with a different scale.

The position of the asymptotic distribution $L(F)$ tends toward $\alpha(F)$, the lower bound of the distribution of one unit. Both γ_F and $\alpha(F)$ reflect properties of the distribution for one unit and are independent of the number of competing units n. On the other hand, the parameter $b_{F,n}$ depends directly on the number of competitors because it reflects the scale of the distribution of minima: the larger the number of competitors, the smaller the scale of the distribution becomes. As n increases to ∞, the scale shrinks toward zero ($b_{F,n} = 0$) and the distribution becomes a step function. A more formal definition of $b_{F,n}$ will be presented in Section 3.

C_1 and C_2 are necessary and sufficient criteria for the existence of an asymptotic distribution (we will also use the term "attractor" distribution) and for the form of the asymptotic distribution to be Weibull. The asymptotic form occurs as $n \to \infty$, but for psychological applications $n = 100$ is a large enough number of competing units that the distribution will be extremely close to the asymptotic form (in some cases much smaller values of n will suffice—this issue will be discussed again in Section 3).

Evaluating the Shape γ_F of an Observed Distribution of Minima

The parameter γ_F can be obtained by solving C_2 (or C_{2a}) if the distribution function F is known or if it can be estimated using a Weibull–linear plot (Weibull, 1951): Let's divide the data into N categories of equal length, C_j, $j = 1, \ldots, N$, and let F_j be the cumulative number of observations in the jth category. The plot of

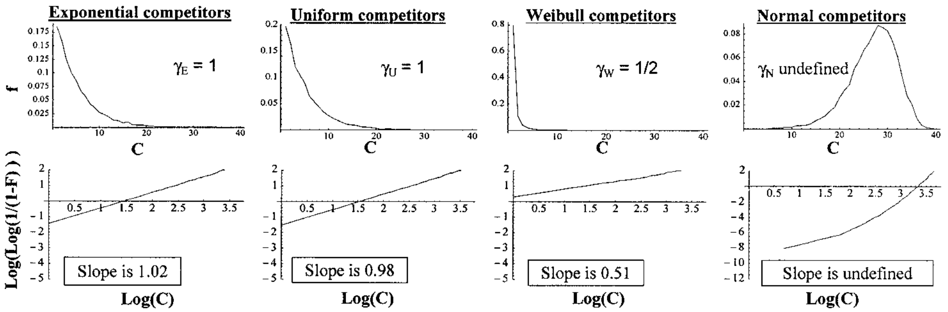

FIG. 3. Distributions of 10,000 minima sampled from 100 independent and identically distributed competitors. The top row presents the relative frequency f obtained, represented as a function of the category number C. Categories are separated into 40 bins of equal size. The bottom row presents the Weibull–linear plot, using a transformation of the cumulative frequency F as a function of a transformation of the category number C. The first column involves exponentially distributed competitors with rate parameter $1/10$. The second column involves uniformly distributed competitors in the range 0 to 10. The third column involves Weibull-distributed competitors with a shape parameter of $1/2$, and a scale parameter of 10. The fourth column involves normally distributed a competitors with a mean of 10 and a standard deviation of 1. The first three distributions of competitors have a constant position at zero.

$$\mathrm{Log}\left(\mathrm{Log}\left(\frac{1}{1-F_j}\right)\right) \quad \text{vs.} \quad \mathrm{Log}(C_j)$$

will yield a straight line with slope γ_F.[4]

As an illustration, in Fig. 3, we generated 10,000 minima sampled from $n = 100$ competitors. The figure shows both the probability density plot and the Weibull–linear plot using exponentially distributed units ($\gamma_E = 1$), uniformly distributed units ($\gamma_U = 1$), and Weibull–distributed units with shape $\gamma_W = 1/2$. These three distributions satisfy C_1 and C_{2a} (and therefore C_2). The γ values have been obtained by solving C_2 and are in good agreement with the slopes of the Weibull linear plots. We also sampled minima from normally distributed units (violating both C_1 and C_2.) As seen in the last plot of Fig. 3, the Weibull–linear plot is no longer linear. To generate samples, we used Mathematica (Wolfram, 1996) and the program found in Listing 1.

1. VARIABILITY IN POSITION

We introduce variability in position by replacing the position parameter $\alpha(F)$ with a random position **Y**. Equivalently, we add to each sample from the unit distribution **X** a sample from a distribution **Y**. For each value sampled from **Y**, we may think of the unit distribution **X** as being shifted along the horizontal axis by that amount, thus achieving position variability.

[4] Categories (or bins) extend from the smallest to the largest sampled values with intervals of equal length (not equal area, as with quantiles). Using category numbers in the Weibull linear plot as we did in the figures is only a matter of convenience, and by doing so the lower bound $\alpha(F)$ and the scale $b_{F,n}$ of the distribution of minima are lost. The figures all start at category 1 and extend up to the last category, Category 40, in all the Weibull linear plot. However, the shape (the main concern in this paper) is preserved.

LISTING 1

Generating Minima from a Uniform Distribution (with Range 0...10) using Mathematica. The Number of Competitors is 100, the Sample Size is 10,000, and the Data Are Plotted Using 40 Categories.

```
<< Statistics‘ContinuousDistributions’
<< Statistics‘DataManipulation’

OneTrial:=Min[Table[Random[UniformDistribution[0,10]], {100}]];
OneSession=Table[OneTrial,{10000}];

freq=BinCounts[
  OneSession,{Min[OneSession],
    Max[OneSession],(Max[OneSession]-Min[OneSession])/40}];
ListPlot[freq, PlotJoined → True];

cumfreq=CumulativeSums[freq];
ListPlot[cumfreq, PlotJoined → True];
```

Let **X** and **Y** be independent random variables. We want to describe the behavior of $(\mathbf{X}+\mathbf{Y})_{1:n} := \min(\mathbf{X}_1+\mathbf{Y}_1, \mathbf{X}_2+\mathbf{Y}_2, \ldots, \mathbf{X}_n+\mathbf{Y}_n)$. Hence, the problem of finding the asymptotic distribution of the minimum of a random variable having a random position parameter is to find the asymptotic distribution of the minimum of a convolution. The convolution is denoted hereafter by the operator *: $F * G$ is the convolution of F and G.

THEOREM 1. *Let $F(x)$ and $G(x)$ be distribution functions with finite lower bounds $\alpha(F)$ and $\alpha(G)$, respectively. Suppose further that F and G satisfy C_2 (Eq. 0.2) with parameters γ_F and γ_G. Then $F * G(x)$ has a lower bound $\alpha(F * G) = \alpha(F) + \alpha(G)$ and satisfies C_2 with $\gamma_{F*G} = \gamma_F + \gamma_G$.*

It is clear that the lower bound of the convolution must be the sum of the two lower bounds since the convolution represents a sum. Proving the theorem therefore requires proving the assertion concerning C_2. Fortunately, Theorem 1 has already been proven. In fact, it is well known that the product of two regularly varying functions is regularly varying as well (see de Haan, 1990; Bingham, Goldie, & Teugels, 1987; Geluk, 1994, 1996).

Perhaps, contrary to intuition, the shape parameter of the convolution does not depend on the relative magnitude (or scale) of both components F and G, but only on their respective shape parameters γ_F and γ_G.

We used a simulation to illustrate the characteristics of the minimum for convolutions of several types of distributions. Using a program like the one found in Listing 2, we generated minima from sums of random samples. Figure 4 presents the probability density plot and the Weibull linear plot from samples obtained with various combinations of distributions. As can be seen, when both distributions satisfy C_1 and C_2, the resulting distribution of the minimum conforms well to a Weibull distribution, and the slope of the Weibull–linear plot is well predicted by the sum of each distribution's γ, as Theorem 1 states. We also plot in Fig. 4 the minima from a convolution of a normal and a uniform distribution. As seen, the resulting Weibull–linear plot is not linear, showing that the attractor, if one exists, is not of Type III.

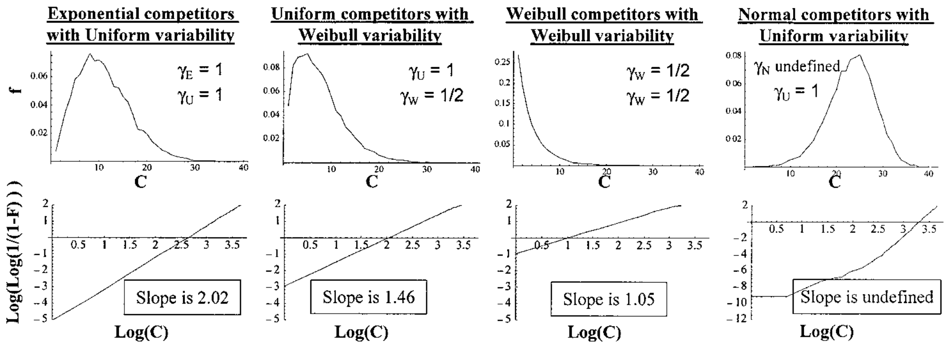

FIG. 4. Distributions of 10,000 minima sampled from 100 independent competitors with variable position parameter. The top row presents the relative frequency f obtained, represented as a function of the category number C. Categories are separated into 40 bins of equal size. The bottom row presents the Weibull–linear plot, using a transformation of the cumulative frequency F as a function of a transformation of the category number C. The first column involves exponentially distributed competitors with a rate parameter of $1/10$ and a position parameter uniformly distributed in the range from 0 to 10. The second column involves uniformly distributed competitors in the range from 0 to 10 with the position parameter Weibull distributed with a shape parameter of $1/2$ and a scale parameter of 10. The third column involves Weibull-distributed competitors with a shape parameter of $1/2$ and a scale parameter of 10, and the position parameter also Weibull distributed with a shape parameter of $1/2$ and a scale parameter of 10. The fourth column involves normally distributed competitors with a mean of 10 and a standard deviation of 1 added to a uniformly distributed position in the range from 0 to 10.

Note that convolution is a symmetrical operation; it makes no difference whether F is thought of as the unit distribution and G the variability in position, or vice versa. Note also that the theorem applies to the convolution of two random variables, but through induction can be extended to the convolution of an arbitrarily large number of random variables.

2. VARIABILITY IN SCALE

Another way to add variability is to make the scale of the units that are racing variable. In such a case, some competitors would be obtained from narrow distributions while others would be sampled from wider distributions. One convenient way to achieve this goal is to generate pairs of samples $(\mathbf{X}_i, \mathbf{Y}_i)$ from independent distributions, say F and G, and to multiply the values of each pair. The minimum of n such samples, $(\mathbf{X} \cdot \mathbf{Y})_{1:n}$, is then given by $\min(\mathbf{X}_1 \cdot \mathbf{Y}_1, \mathbf{X}_2 \cdot \mathbf{Y}_2, ..., \mathbf{X}_n \cdot \mathbf{Y}_n)$.

Products of samples are not commonly studied, and we are not aware of a standard terminology for this operation. In what follows, we will use the term "production" for distributions resulting from the product of two random variables, denoted by the sign $\wedge$ (caret). For example, $F \wedge G(x)$ is called the production of F and G and denotes the distribution of the sample products $\mathbf{X}_i \cdot \mathbf{Y}_i$.

In this section, we will assume that the distributions cannot attain negative values (i.e., $\alpha(F) \geqslant 0$ and $\alpha(G) \geqslant 0$). This is necessary because: (i) scale is an unsigned value and (ii) multiplying a distribution by -1 creates a mirror image distribution, which may no longer be lower bounded.

The following theorems on productions fall into three classes, depending on whether zero is included in the domain of the distribution functions. The proofs can be found in the Appendix.

LISTING 2

Generating Minima from a Convolution of a Uniform Distribution (with Range 0…10) and an Exponential Distribution (with Rate 1/10) Using Mathematica. The Number of Competitors Is 100, the Sample Size Is 10,000, and the Data Are Plotted Using 40 Categories.

```
<< Statistics`ContinuousDistributions`
<< Statistics`DataManipulation`

OneTrial:=
  Min[Table[
    Random[UniformDistribution[0,10]]+
      Random[ExponentialDistribution[1/10]],
      {100}]];
OneSession=Table[OneTrial,{10000}];

freq=BinCounts [
  OneSession, {Min[OneSession] ,
    Max[OneSession],(Max[OneSession]-Min[OneSession])/40}];
ListPlot[freq, PlotJoined → True];

cumfreq=CumulativeSums[freq];
ListPlot[cumfreq, PlotJoined → True];
```

THEOREM 2a. *Let $F(x)$ and $G(x)$ be distribution functions whose lower bounds $\alpha(F)$ and $\alpha(G)$ are equal to 1. Suppose further that F and G satisfy C_2 (Eq. 0.2) with parameters γ_F and γ_G. If* **X** *and* **Y** *denote independent random variables whose distributions are F and G, respectively, then the distribution of the random variable $\mathbf{X}\cdot\mathbf{Y}$, given by $F \wedge G(x)$, satisfies C_2 (Eq. 0.2) with $\gamma_{F\wedge G} = \gamma_F + \gamma_G$ and $\alpha(F \wedge G) = \alpha(F)\cdot\alpha(G) = 1$.*

Although the theorem is formulated strictly in terms of $\alpha(F) = \alpha(G) = 1$, it can be generalized to any positive lower bounds by noting that

$$\min(\mathbf{X}_i \cdot \mathbf{Y}_i) = \alpha(F)\cdot\alpha(G)\cdot\min\,(\mathbf{X}_i/\alpha(F)\cdot\mathbf{Y}_i/\alpha(G)),$$

in which $\mathbf{X}_i/\alpha(F)$ and $\mathbf{Y}_i/\alpha(G)$ are lower bounded at 1. It is worth mentioning that multiplying distribution functions with constants has no effect on the asymptotic shape.

The next theorem applies if one of the random variables includes zero.

THEOREM 2b. *Let $F(x)$ and $G(x)$ be distribution functions with lower bounds $\alpha(F) = 0$ and $\alpha(G) = 1$. Suppose further that F satisfies C_2 (Eq. 0.2) with parameter γ_F. If* **X** *and* **Y** *denote independent random variables whose distributions are equal to F and G, respectively, then the distribution of the random variable $\mathbf{X}\cdot\mathbf{Y}$, denoted by $F \wedge G(x)$, satisfies C_2 (Eq. 0.2) with $\alpha(F \wedge G) = 0$ and $\gamma_{F\wedge G} = \gamma_F$.*

Note. There is no assumption concerning **Y** other than $\mathbf{Y} \geqslant \alpha(G) = 1$.

Theorem 2b is a generalization of an earlier finding by E. Dzhafarov (reported in Logan, 1992). In his proof, F was restricted to be a Weibull distribution. By contrast, the above theorem generalizes to any F satisfying the criterion C_1 and C_2. Again, since $\min(\mathbf{X}_i \cdot \mathbf{Y}_i) = \alpha(G)\cdot\min\,(\mathbf{X}_i \cdot \mathbf{Y}_i/\alpha(G))$, we can generalize to any positive lower bound $\alpha(G)$.

The last demonstration is concerned with cases where both distributions includes zero ($\alpha(F) = \alpha(G) = 0$). At this time, we haven't been able to come up with a solution based on the necessary and sufficient criterion C_2. The following proof assumes the weaker criteria C_{2a} and thus is labeled a proposition. We suspect that a solution based on the necessary and sufficient criterion C_2 is possible, but further work is required. Proposition 2c is reported because many distributions used in psychology satisfy C_{2a} (e.g., the uniform distribution, the Gamma distribution, the exponential distribution, and the Weibull distribution).

PROPOSITION 2c. *Let $F(x)$ and $G(x)$ be distribution functions with zero lower bounds. Suppose further that F and G satisfy criterion C_{2a} (Eq. 0.3) with parameters γ_F and γ_G, respectively. If* **X** *and* **Y** *denote independent random variables whose distributions are equal to F and G, respectively, then*

$$\lim_{t \downarrow 0} \frac{\Pr\{X \cdot Y \leqslant t\}}{t^{\min(\gamma_F, \gamma_G)}} = K \qquad \textit{in the case} \quad \gamma_F \neq \gamma_G,$$

$$\lim_{t \downarrow 0} \frac{\Pr\{X \cdot Y \leqslant t\}}{t^{\gamma_F} \log(1/t)} = K \qquad \textit{in the case} \quad \gamma_F = \gamma_G,$$

Therefore, the distribution of **X** $\cdot$ **Y**, *denoted $F \wedge G(x)$, is regularly varying at 0 and, by the definition of regular variations (Feller, 1966), satisfies C_2 (Eq. 0.2) with $\gamma_{F \wedge G} = \min(\gamma_F, \gamma_G)$.*

The propositions and theorems of Sections 1 and 2 cover many forms of parameter variability and show that the attractor is still a Weibull. Further, it show that the shape γ of the resulting distribution of minima can be predicted with the use of simple arithmetic. However, these asymptotic theories apply when the number of competitors goes to ∞. In the next section, we address more empirical issues.

3. RATE OF CONVERGENCE, DEGENERATION AS n INCREASES, AND APPLICATIONS TO PROCESSING MODELS

We have shown that the distribution of the race winner approaches a Weibull shape as the number of racers increases, under general conditions. Observed response-time distributions sometimes approximate a Weibull, raising the possibility that a race produces such a result. In any event, race models have sometimes been employed in processing models, and it is natural to ask whether the present results will be helpful in the analysis and generation of models.

There are several issues that must be addressed before one can connect the present theoretical results to observed data. We discuss two of them in the following.

3.1. Rate of Convergence

For some theories (e.g., instance theories), the number of units n increases with training. In these cases, performance early in training is determined by a small number of units. The results presented above are concerned with large n, but make no explicit statement on how large n must be for these results to hold.

It is generally assumed as a rule of thumb that n must be larger than 400 for any minimum sampled from distributions satisfying C_2 to be Weibull distributed. In the following, we will show examples where n does not have to be that large before the Weibull distribution is observed.

In Figs. 3 and 4, we showed actual distributions of minima for some unit distributions satisfying C_{2a} (Eq. 0.3) and also some of their convolutions (similar graphs were obtained with production). For all these samples, we used a number of competing units $n = 100$. As seen from the Weibull linear plot, the shape obtained (as measured by the slope) conforms pretty well (although no statistical test is used here) to the expected shape of the attractor. Therefore, as far as the criterion C_{2a} is concerned, $n = 100$ seems to be a large enough number of competitors.

To have a closer look at the evolution of the attractor as n increases, we focus on two specific functions. First, the convolution $F_1 * F_2$, where F_1 is exponential with mean 100 and F_2 is uniform with limits 0 to 200. We generated random samples of 10,000 minima using n units, where n varied between 1 and 128. The left column of Fig. 5 shows the results. The top panel illustrates the probability density function (PDF) for various n. When $n = 1$ (there is no race since there is only one unit), the shape looks like an exponential with an unexpected left-end tail, due to the uniform component of the convolution. This shape is not Weibull, as seen from the bottom panel, where the corresponding line is not linear. With $n = 2$, the shape is close to Weibull, but still the right tail is too low and the deviation from linearity is important, as seen in the bottom plot. However, with 4 units in competition, the attractor does not differ from the expected Weibull with a shape of 2. In this precise situation, therefore, n can be very small and still allow observation of the asymptotic distribution of the minima.

Second, we studied the production $F_1 \wedge F_2$, where F_1 and F_2 are defined as in the previous example. In order to have a shape equal to 2, we added a constant to both samples so that the lower bounds are not zero (Theorem 2a). The right top panel of Fig. 5 shows the PDF for various n. For small n, the same basic pattern of results as above holds. Further, one can see that the rate of convergence is slower: at $n = 4$ and $n = 8$, the distribution of minima still has not achieved its asymptotic form (slope smaller than 2 and a small deviation from linearity respectively). However, for $n = 16$, the attractor shape is reached. Basically, production has a slower rate of convergence because multiplication can create much larger deviations in the unit distributions than can addition. Yet, the slow down in convergence is only in the order of 1 magnitude.

Finally, as an extreme, $n = 1$ can suffice when minima are sampled from a Weibull distribution (including the exponential as a special case). This is so because this distribution is itself its own attractor.

3.2. *Degeneration as n Increases*

In the previous theorems, the scale factor $b_{F,n}$ is used to keep the scale of the distribution of the race winner approximately constant as n increases. Without this normalizing factor, the larger the number of competing units, the smaller is the scale of the distribution of the minimum. In the limit, the scale drops to zero. This degeneration of $L(F)$ is unavoidable no matter what type of distribution F is assumed and is only a question of having an n large enough (Colonius, 1995).[5] As the asymptotic distribution approaches a step function, it contributes none of the variance of the observed response-time distribution and becomes useless as an explanatory tool whatever its shape. From a purely mathematical viewpoint one could argue that these scaling issues are irrelevant: No matter how much shrinkage has occurred at a given n one could assume a sufficiently large scale for the unit distributions that the resultant scale for the minimum at n matches the observed variance. For example, if one assumes that 10^7 memory traces race when a memory probe occurs and that the response time distribution observed is roughly a Weibull with a standard deviation of 300 ms, this may be consistent with a unit distribution having a standard deviation of 10 h. From a psychological viewpoint, however, such a large scale for a unit distribution could well be implausible and unjustifiable. Thus for situations in which a large number of units are racing, we must ask in what cases a plausible scale of the unit distribution could give rise to a scale of the minimum distribution that is large enough to capture a significant portion of the variance.

It turns out that even for large n there are cases in which the scale of the minima can be in the same general ballpark as the scale of the unit distributions. To illustrate this point, we first define $b_{F,n}$ more formally, then we present some numerical estimations of $b_{F,n}$ for productions and convolutions.

DEFINITION OF $b_{F,n}$. The asymptotic distribution is usually presented with the normalizing constants $\alpha(F)$ and $b_{F,n}$ as a function of

$$\frac{X_{1:n}-\alpha(F)}{b_{F,n}},$$

where $\alpha(F)$ is defined in Eq. (0.1). Galambos (1978) defines the sequence $\{b_{F,n}\}$ as

$$b_{F,n} := \sup\{x \mid F(x) \leqslant 1/n\} - \alpha(F). \tag{3.1}$$

In other words, $b_{F,n}$ is the point that cuts off the first n-tile of the distribution of the minimum of n units. We note that

[5] Another question that relates to both Sections 3.1 and 3.2 is whether the asymptotic shape is achieved before degeneration. The following argument, kindly provided by E. Dzhafarov, shows that the asymptotic shape is obtained infinitely faster than degeneration is. Let $F(x)$ be a continuous distribution function on $x>0$ satisfying C_1 and C_2. We can find coefficients a_n, b_n such that $F(a_n+b_nx)^n \to W(x)$ as $n\to\infty$, where $W(x)$ is the Weibull distribution. This convergence is uniform; i.e., $\sup_x\{F(a_n+b_nx)^n - W(x)\} \to 0$ as $n\to\infty$. As we know also, $F(x)^n \to H(x)$, where $H(x)$ is the standard Heaviside function (jumping at $x=0$ from zero to one). This convergence is pointwise, but not uniform: $\sup_x\{F(x)^n - H(x)\} = 1$ for any n. As a result,

$$\frac{\sup_x \{F(a_n+b_nx)^n - W(x)\}}{\sup_x \{F(x)^n - H(x)\}} \to 0 \qquad \text{as} \quad n\to\infty.$$

This argument applies to the convergence of minima using $1-[1-F(x)]^n$ and is also valid for type-I and type-II distributions of extremes.

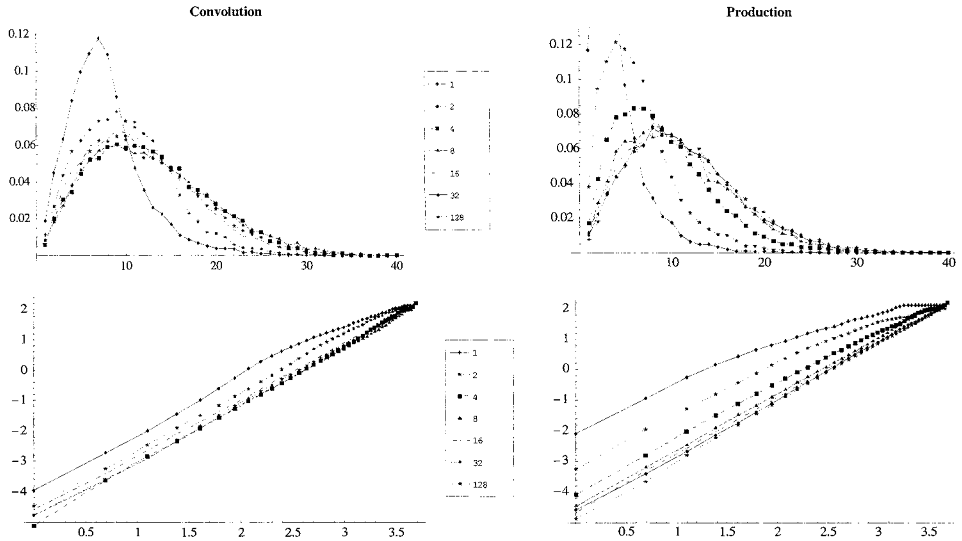

FIG. 5. Illustration of minima sampled from $F_1 * F_2$ where F_1 is exponential with mean 100 and F_2 is uniform in the limits 0 to 200 as a function of the number of competing units n. Sample size is 10,000. Top panel shows observed probability density functions as a function of category number C; bottom panel shows Weibull linear plot of the above.

$$\Pr\left\{\frac{\mathbf{X}_{1:n}-\alpha(F)}{b_{F,n}} \leqslant x\right\} = \Pr\{\mathbf{X}_{1:n} \leqslant \alpha(F)+b_{F,n}x\}$$

and, by substitution into Eq. (0.4);

$$L(F)\,(\alpha(F)+b_{F,n}x) = W(\gamma_F, 0, 1)$$

is a normalized Weibull with position zero and scale 1. Therefore $b_{F,n}$ is a reasonable normalization factor that keeps the scale approximately constant as n increases.

In the general case where F satisfies the criteria C_1 and C_2, it is difficult to obtain a general expression for $b_{F,n}$. However, a general formula exists for distribution functions satisfying C_1 and C_{2a}, that is, for functions that are exactly the power curve in the left-end tail. From the continuous assumption, we can equate $F(b_{F,n})$ to $1/n$ (Eq. 3.1) and solve for $b_{F,n}$ using C_{2a}. The result is given by

$$b_{F,n} = (K_F n)^{-1/\gamma_F}, \tag{3.2}$$

where $K_F > 0$ is the constant defined in Eq. (0.3) for the original distribution F. What Eq. (3.2) says is that $b_{F,n}$ is a negatively accelerated power curve since the exponent is negative. As such, $b_{F,n}$ is a slowly reducing curve with zero asymptote: if $b_{F,n}$ decreases by a factor Δ when the first n units are added to the race, it will take n^2 units for $b_{F,n}$ to reduce by a factor Δ again (Newell and Rosenbloom, 1981).

Numerical Estimation of $b_{F,n}$. In this section, we study the behavior of $b_{F,n}$ as n increases. The parameter values used are arbitrary, but scales can be thought of as being expressed in milliseconds.

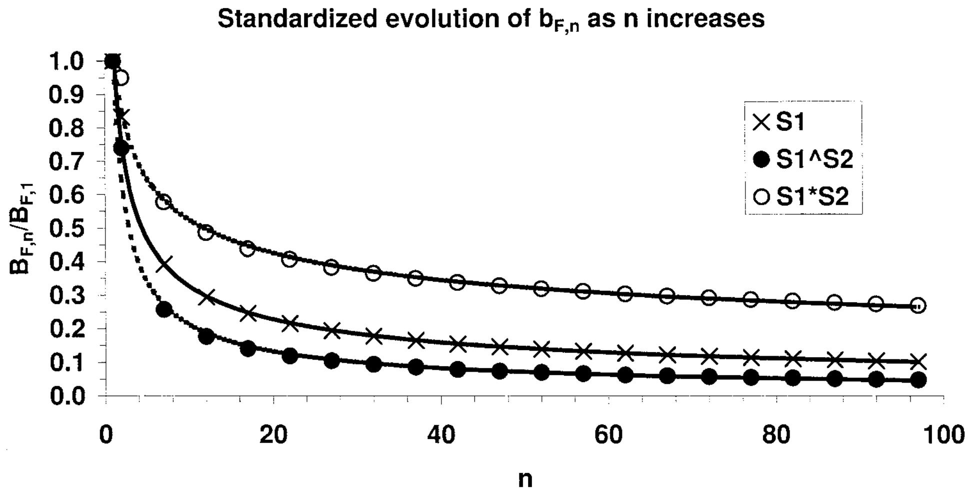

FIG. 6. Evolution of $b_{F,n}$ as the number of competing units n is increased from 1 up to 100. Two distributions, S_1 and S_2, and the convolution and the production of these two are shown. S_1 has a scale $b_{S_1,1}$ of 100 while S_2 has a scale $b_{S_2,1}$ of 200. All the points have been estimated using Eq. (3.2) and numerical integration techniques when the distribution function was not available in closed form ($S_1 \wedge S_2$ and $S_1 * S_2$). The solid lines represent exact analytical solutions (in the case of S_1 and S_2 only); the dashed lines are power curves estimated through best-fitting procedures (in the case of $S_1 \wedge S_2$ and $S_1 * S_2$ only).

In the first run, units have a distribution function S_1 being a Weibull distribution (with scale parameter $b_{S_1,1} = 200$, shape $\gamma_{S_1} = 2$, and lower bound $\alpha(S_1) = 0$). With 100 units in competition, the scale of the distribution of minima $b_{S_1,100}$ drops to less than 20, a 90% reduction. Such a drop is difficult to reconcile with psychological plausibility. In a second simulation, we used a function S_2 similar to S_1 but with the scale parameter $b_{S_2,n}$ doubled to 400. Results did not change much since the scale of the distribution of minima $b_{S_2,100}$ is now around 40, exactly the same 90% drop, since the two curves have the same shape.

In both simulations, $b_{S_1,n}$ and $b_{S_2,n}$ were estimated by computing Eq. (3.1) using minimization techniques. Fig. 6 illustrates the relative behavior of $b_{S_1,n}$ for n ranging from 1 to 100 when the starting position of the curve $b_{F,1}$ is discounted. In other words, the figure shows the proportion of reduction in scale when n increases. $b_{S_2,n}$ is not shown since it is identical to $b_{S_1,n}$. The full line shows the exact solution obtained by computing Eq. (3.2). An exact solution is possible since K_{S_1} and K_{S_2} are given after solving Eq. (0.3). In the case of Weibull distributions, $K_W = (1/b_{W,1})^{\gamma_W}$, that is, the reciprocal of the scale, raised to the shape parameter of the distribution. With 100 units, only 10% of the original scale is left. With 1000, 3% remains. If $b_{S_1,1}$ is the standard deviation when one unit is present, the same system with 1000 units would have a standard deviation of 6 ms.

Let us consider next what happens if we retain S_2 but add variability in scale or position using S_1. We therefore consider $S_1 \wedge S_2$ and $S_1 * S_2$.[6] Fig. 6 also illustrates $b_{S_2 * S_2,n}$, and $b_{S_2 \wedge S_2,n}$, for n ranging from 1 to 100. The upper curve of Fig. 6 shows

[6] A preliminary proposition not published showed that if both S_1 and S_2 satisfy C_{2a}, then the convolution $S_1 * S_2$ will also. Conversely, from Proposition 2c, in the case where $\gamma_F = \gamma_G$, we see that even if S_1 and S_2 satisfy C_{2a} the shape of the left-end tail of $S_1 \wedge S_2$ is not a power curve, but the product of a slowly-varying function and a power curve. Therefore, $S_1 * S_2$ satisfies C_{2a} but not $S_1 \wedge S_2$.

the scale for the distribution of minima for $S_1 * S_2$ while the lower one shows $S_1 \wedge S_2$. Dots on these curves could not be solved analytically since these operations do not yield closed-formed solutions and therefore have been estimated using numerical integration techniques.

The convolution decreases much more slowly toward zero. For 100 units, the scale is ca. 25% of the scale of the original distribution. After 1000 units with variability in position are added, the scale $b_{S_2 * S_2, 1000}$ is still 14% of the starting scale, around 60 ms for a starting scale of 400 ms, five times larger than when no variability is present.

The production $S_1 \wedge S_2$ decreases faster than S_1 or S_2 taken individually.

The dotted lines represent estimated power curves. These curves show, as an approximation, that first, a power curve describes well the behavior of $b_{S_2 * S_2, n}$, but less effectively $S_1 \wedge S_2$ which does not satisfy the weaker criterion C_{2a} used in the first place to derive the relation found in Eq. (3.2). Close inspection of Proposition 2c suggests that

$$b_{F,n} = (K_F n \ln n)^{-1/\gamma_F}$$

is a better description of the reduction in the scale.

Second, convenient constants to parametrize the curves are given by

$$K_{S_1 * S_2} \approx K_{S_1} + K_{S_2},$$
$$K_{S_1 \wedge S_2} \approx K_{S_1} \cdot K_{S_2}.$$

Therefore, in the case of convolutions, the initial scale is the sum of the individual scales and the rate is $-1/(\gamma_F + \gamma_G)$, that is, a higher starting point and a smaller rate of decrease.

The point of Fig. 7 to remember is that even though scales reduce and will ultimately reach zero in all cases, they can do so slowly for $L(F * G)$. In fact, with 100,000 units, the scale is still more than 4% the size of the original scale, a figure achieved with only 100 units when no variability is introduced!

4. SUMMARY AND DISCUSSION

In this text, we have shown that the identically distributed assumption is not necessary in order to draw conclusions about the asymptotic distribution of the minima. We described two ways to have nonidentical distributions: one is to affect the lower bound of the competing unit's distributions, the other is to affect the scale, or the range, of the competing distributions. As an analogy with a race, we could say in the first case that the runners don't start at the same moment, while the second case corresponds to runners racing on different road surfaces.

We summarize in Table 1 the various theorems that predict the asymptotic distribution of minima when the conditions are satisfied.

As can be deduced from Table 1, we see that convolution is a closed operator with respect to the minima of Type III: whatever the distributions F and G, if each

TABLE 1

Asymptotic Distribution of Minima when Convolution or Production Are Applied to Two Distribution Functions *F* and *G*

Operation	Condition to be satisfied	Asymptotic distribution of minima	Solved in
F	C_1, C_2	$L(F) = W(\gamma_F, \alpha(F), b_{F,n})$	
$F * G$	C_1, C_2	$L(F * G) = W(\gamma_F + \gamma_G, \alpha(F) + \alpha(G), b_{F \wedge G, n})$	Theorem 1
$F \wedge G$			
$\alpha(F) > 0, \alpha(G) > 0$	C_1, C_2	$L(F \wedge G) = W(\gamma_F + \gamma_G, \alpha(F) \cdot \alpha(G), b_{F \wedge G, n})$	Theorem 2a
$\alpha(F) = 0, \alpha(G) > 0$	C_1, C_2 on F only	$L(F \wedge G) = W(\gamma_F, 0, b_{F,n})^a$	Theorem 2b
$\alpha(F) = \alpha(G) = 0$	C_1, C_{2a}	$L(F \wedge G) = W(\text{Min}(\gamma_F, \gamma_G), 0, b_{F \wedge G, n})$	Proposition 2c

Note. The criteria C_{2a} is included in C_2; L denotes the asymptotic distribution of minima and W the Weibull distribution.

[a] Generalization of a special case studied by E. Dzhafarov (reported in Logan, 1992).

individually has minima of Weibull type, then their convolutions do as well (for more on the closure properties of regularly varying functions, see Embrechts & Goldie, 1980; Bingham *et al.*, 1987; Geluk, 1994, 1996). Further, this operation has commutativity $(L(F * G) = L(G * F))$ and associativity $(L((F * G) * H) = L(F * (G * H)))$ properties. In this respect, any number of convolved distributions will remain in the domain of attraction of their elements. The only drawback concerns empirical issues: The shape of convolved distributions can only increase. For example, convolving three exponential distributions yields an asymptotic distribution of Type III with shape $\gamma = 3$. However, empirical measures of the shape of reaction times find typical values around 2 (for example, see Logan, 1992, and Cousineau, in preparation). If position variability is the only source of perturbation in the race, then either there is a very limited number of convolved components or each contributes a very small value to the overall shape. Fortunately, the other operation studied, production, introduces some nonlinearity into the shape.

Production describes the distribution obtained from the product of two samples. For example, production could be used to describe races where the response's threshold is a random variable. Because we haven't found a general solution to the last case (Proposition 2c, see Table 1), we can only conjecture that production is closed with respect to the minima of Type III. However, since distribution functions satisfying the criteria C_{2a} also fulfill the more general criteria C_2, we know production to be closed with respect to the minima of Type III for some of the most often used distribution functions in psychology (e.g., the exponential distribution and its related distributions). Since production has also commutativity and associativity properties, results can be generalized to any number of distributions.

Further, productions can be mixed with convolutions in various ways without affecting the results. However, distributivity is not a property of these operators, hence $L(F * (G \wedge H)) \neq L(F * G \wedge F * H)$.

Finally, we addressed the question of degenerability. For theories where n increases with practice, the scale $b_{F,n}$ of the observed distribution of minima decreases and, past some large n, might look like a step function. Although we cannot eliminate degeneration, we showed that with convolution and production n might have to be so large that it is implausible to assume that number of units, even if these units without parameter variability can degenerate rapidly. This makes it possible that this phenomenon will never be observed.

The above results strike a decisive blow to the identically distributed assumption. A very general form of variability in position or scale can be introduced without affecting the existence of an asymptotic distribution of minima. Therefore, the random number generator most frequently found on computers (which mimics the uniform distribution) can be put in the place of the position parameter or scale parameter for convenient, easy-to-do simulations.

Galambos (1978) followed the same line of thought and challenged the independent assumption. Instead, he assumed that the units are located inside of some space so that the distance between them can be computed. He showed that, as long as the dependencies between units weaken with distance, the asymptotic distribution of extreme remains unchanged (as long as C_1 and C_2 are true, of course). This assumption is plausible if we assume that dependency (e.g., inhibition) is a signal requiring some time to travel from one unit to the other: if a large distance separates two units, the signal may well not arrive in time. Consistent with race models, time is a crucial factor.

From the above, it is clear that γ is not simply a free parameter but reflects some fundamental aspect of the units. Compatible with this claim, we showed elsewhere (Cousineau and Larochelle, in preparation) that the empirical shape parameter seems to be a constant over practice in a visual and memory search task. Further, empirical results provide strong constraints since the shape parameter is seldom found to be larger than 2 (Logan, 1992).

With the various points made here, there remains no in-principle reason to reject race models. They can accommodate empirical distributions of reaction times (nondegenerate functions with shape typically around 2). Also, they can do so with plausible competitors, i.e., competitors that are not identically distributed. For example, noise can be introduced in the starting position without affecting the predictions on the nature of the observed reaction times. However, the theorems impose limits on the nature of the distributions used. For example, we saw that the convolution of a uniform distribution and a normal (Gaussian) distribution is not a member of the Type-III attractor (see Fig. 4). Only distributions with a power curve left-end tail (criterion C_2, Eq. 0.2) can be manipulated. Therefore, in the context of race models (and maybe for psychological models in general), normally distributed noise is far from the simplest choice available to modelers.

Race models provide a nice and intuitive account of performance. Further, some models implement learning by assuming that each experience with a stimulus results in one more trace being added to the race. As we have shown in Section 3.2, the net result is a decrease in the scale of the distribution of minima. One question remains, though: On a given trial, which traces are included in the race? Does the system restrict itself to traces identical to the stimulus, or is there a gradation in the traces recruited? This selection problem has been answered in various ways in the past: Logan (1988) assumed a strict selection, stating that only traces identical to the probe were to participate in the race. On the other hand, Bundesen (1990) assumed no selection at all, but more closely similar traces had a greater chance of winning the race by having a smaller scale. If selection is part of a model (as a preprocessor, for example), it is important to characterize its behavior, since its distribution will convolve with the distribution of the winner. A unified model not only must describe what information flows from the input to the output, it must also consider *when* such outputs are made. This signifies a totally different approach to network modeling.

Finally, Colonius (1995) asked the question "Why the Weibull?" in the context of the instance-based theory of automatization (Logan, 1992, 1995). The answer is that it does not have to be a Weibull. Uniform and Gamma distributions are as good contenders since position or scale variability, even of small magnitude, can alter profoundly the shape of the distribution. Some simulations showed that simple convolutions converge very rapidly toward a Weibull distribution. With as few as $n = 8$ competitors, it is difficult to distinguish the distribution of minima from a Weibull distribution.

Parameter variability can have important impact on a model's prediction (and so they were studied by, e.g., Colonius, 1990; Ulrich & Giray, 1986; and Van Zandt & Ratcliff, 1995). We have shown in this text that this is also true for race models, but that these effects can be understood in terms of simple arithmetic.

APPENDIX

Proof of Theorem 2a. Again, the assertion concerning $\alpha(F \wedge G)$ is obvious. We will reduce the product question to one involving the addition of two independent random variables. Let **X** and **Y** be independent random variables with the corresponding distributions F and G.

Define the (independent) random variables

$$\mathbf{U} = \mathbf{X} - 1 \qquad \text{and} \quad \mathbf{V} = \mathbf{Y} - 1.$$

Note that the lower bounds of both **U** and **V** are 0. Also,

$$\begin{aligned} \mathbf{X} \cdot \mathbf{Y} &= (\mathbf{U} + 1)\,(\mathbf{V} + 1) \\ &= 1 + \mathbf{U} + \mathbf{V} + \mathbf{UV}. \end{aligned}$$

In order to prove the theorem, we must show that the distribution function, $H(x)$, of the random variable

$$\mathbf{U} + \mathbf{V} + \mathbf{UV}$$

satisfies C_2 (Eq. 0.2) with the correct γ value. To do this, we will show that $H(x)$ is asymptotically equal to

$$F_U * F_V.$$

That is, as $x \to 0$, $H(x)$ behaves as though it were a convolution of the **U** and **V** distributions.

Step 1. $H(x) < F_U * F_V(x)$.

This is clear since, by definition,

$$\begin{aligned} H(x) &= \Pr\{\mathbf{U}+\mathbf{V}+\mathbf{UV} \leqslant x) \\ &< \Pr\{\mathbf{U}+\mathbf{V} \leqslant x\} \\ &= F_U * F_V(x). \end{aligned}$$

Step 2. $H(x) \geqslant F_U * F_V(\frac{x}{1+x})$.

Suppose that $\mathbf{U}+\mathbf{V}+\mathbf{UV} \leqslant x$. Since both $\mathbf{U}$ and $\mathbf{V}$ are nonnegative, this implies that $\mathbf{U} \leqslant x$ and $\mathbf{V} \leqslant x$. It follows that

$$\begin{aligned} \mathbf{U}+\mathbf{V}+\mathbf{UV} &\leqslant \mathbf{U}+\mathbf{V}+x\mathbf{V} \\ &< \mathbf{U}+\mathbf{V}+x\mathbf{V}+x\mathbf{U} \\ &= (\mathbf{1}+x)\,\mathbf{U}+(\mathbf{1}+x)\,\mathbf{V} \\ &= (\mathbf{1}+x)\,(\mathbf{U}+\mathbf{V}). \end{aligned}$$

We state this relation in terms of the distribution functions

$$\begin{aligned} H(x) &= \Pr\{\mathbf{U}+\mathbf{V}+\mathbf{UV} \leqslant x\} \\ &\geqslant \Pr\{(1+x)\,(\mathbf{U}+\mathbf{V}) \leqslant x\} \\ &= \Pr\left\{\mathbf{U}+\mathbf{V} \leqslant \frac{x}{1+x}\right\} \\ &= F_U * F_V\left(\frac{x}{1+x}\right). \end{aligned}$$

Step 3. Show that

$$\lim_{x \downarrow 0} \frac{H(x)}{F_U * F_V(x)} = 1.$$

One inequality is clear from Step 1: $H/F_U * F_V \leqslant 1$. This holds in the limit as well. Step 2 provides the other inequality, dividing each side by $F_U * F_V$,

$$\frac{H(x)}{F_U * F_V(x)} \geqslant \frac{F_U * F_V\left(\dfrac{x}{1+x}\right)}{F_U * F_V(x)}.$$

For any fixed $t > 1$, we have $\frac{x}{1+x} > x/t$ if x is sufficiently small. Then, for the purpose of finding a limit, we may state that

$$\frac{H(x)}{F_U * F_V(x)} \geqslant \frac{F_U * F_V(x/t)}{F_U * F_V(x)}.$$

Now, take the limit of the right-hand expression. The Theorem 1 for convolutions states that the limit of the right-hand expression is $t^{-\gamma}$. This gives

$$\lim_{x \downarrow 0} \inf \frac{H(x)}{F_U * F_V(x)} \geqslant t^{-\gamma}.$$

But $t > 1$ is arbitrary so the limit infimum has to be 1. This shows that the limit exists and equals one.

Step 4. Show that

$$\lim_{x \downarrow 0} \frac{H(x/t)}{H(x)} = t^{-\gamma_U - \gamma_V}.$$

The Theorem 1 on convolutions states that this identity holds for the distribution function $F_U * F_V$. Now, we simply repeat the proof of Theorem 1, that is,

$$\frac{H(x/t)}{H(x)} = \frac{H(x/t)}{F_U * F_V(x/t)} \cdot \frac{F_U * F_V(x/t)}{F_U * F_V(x)} \cdot \frac{F_U * F_V(x)}{H(x)},$$

and we take each limit individually to obtain $1 \cdot t^{-\gamma_U - \gamma_V} \cdot 1$ for the limit of the left-hand expression. This proves Theorem 2a. ▮

Proof of Theorem 2b.

$$\Pr\{X \cdot Y \leqslant t\} = \int_1^\infty F(t/y)\, dG(y).$$

We will show that

$$\lim_{t \downarrow 0} \frac{\Pr\{X \cdot Y \leqslant t\}}{F(t)} = K.$$

Since F satisfies C_2, the usual argument shows that the distribution function satisfies C_2. Choose any positive sequence t_n such that $t_n \to 0$. Consider the ratio

$$\frac{\Pr\{X \cdot Y \leqslant t_n\}}{F(t_n)} = \int_1^\infty \frac{F(t_n/y)}{F(t_n)}\, dG(y).$$

Since $y \geqslant 1$ in this integral and F is increasing, for any y and t_n,

$$\frac{F(t_n/y)}{F(t_n)} \leqslant 1.$$

That is, each integrand is dominated by the constant 1. Also, for each fixed y,

$$\lim_{n \to \infty} \frac{F(t_n/y)}{F(t_n)} = y^{-\gamma_F}.$$

This follows from C_2. We may apply the *dominated convergence theorem* (Galambos, 1978, p. 317),

$$\lim_{n\to\infty}\int_1^\infty \frac{F(t_n/y)}{F(t_n)}\,dG(y)=\int_1^\infty \lim_{n\to\infty}\frac{F(t_n/y)}{F(t_n)}\,dG(y)$$
$$=\int_1^\infty y^{-\gamma_F}\,dG(y).$$

The above integral is a constant K, and we have shown that $\lim_{t\downarrow 0}\frac{\Pr\{X\cdot Y\leqslant t\}}{F(t)}=K$. This proves Theorem 2b. ▮

Proof of Proposition 2c.

$$\Pr\{X\cdot Y\leqslant t\}$$
$$=\Pr\{Y>\sqrt{t},X\cdot Y\leqslant t\}+\Pr\{X>\sqrt{t},X\cdot Y\leqslant t\}+\Pr\{X\leqslant\sqrt{t},Y\leqslant\sqrt{t}\}$$

This holds since the condition $x\cdot y\leqslant t$ consists of the three disjoint events described in the equation above.

First, we show that $\Pr\{Y>\sqrt{t},X\cdot Y\leqslant t\}$ is asymptotically equal to $F(t)\int_{\sqrt{t}}^\infty y^{-\gamma_F}\,dG(y)$;

$$\Pr\{Y>\sqrt{t},X\cdot Y\leqslant t\}=\int_{\sqrt{t}}^\infty F(t/y)\,dG(y).$$

Now, if $\sqrt{t}<y<\infty$, then $0<t/y<\sqrt{t}$.

We denote t/y by x. Note that for any y in the integral above we have $0<x<\sqrt{t}$. That is, any choice for x is near zero if t is chosen to be near 0. We now use the hypothesis C_{2a}. It states that

$$F(x)\approx K_F x^{\gamma_F},$$

and this approximation holds uniformly throughout the range $x<\sqrt{t}$. We may replace the integrand $F(x)$ with the expression $K_F(t/y)^{\gamma_F}$ to obtain

$$\Pr\{Y>\sqrt{t},X\cdot Y\leqslant t\}\approx\int_{\sqrt{t}}^\infty K_F(t/y)^{\gamma_F}\,dG(y)$$
$$=K_F t^{\gamma_F}\int_{\sqrt{t}}^\infty y^{-\gamma_F}\,dG(y)$$
$$\approx F(t)\int_{\sqrt{t}}^\infty y^{-\gamma_F}\,dG(y).$$

This completes the first step.

Next, we show that $\int_{\sqrt{t}}^\infty y^{-\gamma_F}\,dG(y)$ has three different limits as $t\downarrow 0$, depending on the relation of γ_F and γ_G.

Case 1. $\gamma_F<\gamma_G$. We integrate by parts,

$$\int_{\sqrt{t}}^\infty y^{-\gamma_F}\,dG(y)=y^{-\gamma_F}G(y)\Big|_{y=\sqrt{t}}^{y=\infty}+\gamma_F\int_{\sqrt{t}}^\infty G(y)\,y^{-\gamma_F-1}\,dy$$
$$=-t^{\gamma_F/2}G(\sqrt{t})+\gamma_F\int_{\sqrt{t}}^\infty G(y)\,y^{-\gamma_F-1}\,dy.$$

Now, since $G(\sqrt{t}\,) \approx K_G t^{\gamma_G/2}$, the first term tends toward zero if $t \to 0$. Also, for y near 0, $G(y)\, y^{-\gamma_F-1} \approx K_G\, y^{\gamma_G-\gamma_F-1}$. Since the exponent is larger than -1, this function has a finite integral over $[0, 1]$. We obtain

$$\lim_{t\downarrow 0} \int_{\sqrt{t}}^{\infty} y^{-\gamma_F}\, dG(y) = \gamma_F \int_0^{\infty} G(y)\, y^{-\gamma_F-1}\, dy = K_4 < \infty.$$

Consequently, we see that

$$\Pr\{Y > \sqrt{t}, X\cdot Y \leqslant t\} \approx K_4 F(t).$$

Case 2. $\gamma_F > \gamma_G$. We still have

$$\int_{\sqrt{t}}^{\infty} y^{-\gamma_F}\, dG(y) = -t^{-\gamma_F/2} G(\sqrt{t}\,) + \gamma_F \int_{\sqrt{t}}^{\infty} G(y)\, y^{-\gamma_F-1}\, dy.$$

However, $t^{-\gamma_F/2} G(\sqrt{t}\,) \approx t^{(\gamma_F-\gamma_G)/2} \to +\infty$ as $t \downarrow 0$.

The integral has the same behavior. If $\varepsilon > 0$,

$$\int_{\sqrt{t}}^{\varepsilon} y^{-\gamma_F-1} G(y)\, dy \approx K_G \int_{\sqrt{t}}^{\varepsilon} y^{\gamma_G-\gamma_F-1}\, dy \approx K_3 t^{(\gamma_F-\gamma_G)/2}.$$

Consequently, in this case,

$$\Pr\{Y > \sqrt{t}, X\cdot Y \leqslant t\} \approx F(t)\, t^{(\gamma_F+\gamma_G)/2} K_4 \approx K_4 t^{(\gamma_F+\gamma_G)/2}.$$

Since $\gamma_F > \gamma_G$, we see that $t^{(\gamma_F+\gamma_G)/2}$ is much smaller than $G(t)$. That is,

$$\lim_{t\downarrow 0} \frac{\Pr\{Y > \sqrt{t}, X\cdot Y \leqslant t\}}{G(t)} = 0.$$

The notation $A \ll B$ means that $\lim_{t\downarrow 0} \frac{A}{B} = 0$ as $t \to 0$. The result above may be restated as $\Pr\{Y > \sqrt{t}, X\cdot Y \leqslant t\} \ll G(t)$.

Case 3. $\gamma_F = \gamma_G$.

$$\begin{aligned}\int_{\sqrt{t}}^{\infty} y^{-\gamma_F}\, dG(y) &\approx -t^{-\gamma_F/2}\cdot K_G \cdot t^{\gamma_F/2} + \gamma_F \int_{\sqrt{t}}^{\varepsilon} K_G\, y^{\gamma_F-\gamma_G-1}\, dy \\ &\approx -K_G + \gamma_F K_2(-1/2 \log(t))\end{aligned}$$

Finally, the third term is given by $\Pr\{X \leqslant \sqrt{t}, Y \leqslant \sqrt{t}\} = F(\sqrt{t}\,)\, G(\sqrt{t}\,)$. To complete the proof, we return to the original decomposition of $\Pr\{\mathbf{X}\cdot\mathbf{Y} \leqslant t\}$.

Now, in Case 1, $\gamma_F < \gamma_G$, the three terms are

$$\begin{aligned}\Pr\{Y > \sqrt{t}, X\cdot Y \leqslant t\} &\approx K_4 F(t),\\ \Pr\{X > \sqrt{t}, X\cdot Y \leqslant t\} &\ll F(t) \qquad \text{(see Case 2)},\\ \Pr\{X \leqslant \sqrt{t}, Y \leqslant \sqrt{t}\} &= F(\sqrt{t}\,)\, G(\sqrt{t}\,) \ll F(t).\end{aligned}$$

This gives the result $\Pr\{\mathbf{X}\cdot\mathbf{Y} \leqslant t\} \approx K_4 F(t)$.

Case 2 is exactly the same. $\Pr\{\mathbf{X}\cdot\mathbf{Y}\leqslant t\}\approx K_4G(t)$.

In Case 3, $\gamma_F=\gamma_G$, we have

$$\Pr\{X>\sqrt{t},\, X\cdot Y\leqslant t\}\approx K_4F(t)\log(1/t),$$

$$\Pr\{Y>\sqrt{t},\, X\cdot Y\leqslant t\}\approx K_5G(t)\log(1/t),$$

$$\text{and}\quad \Pr\{X\leqslant\sqrt{t},\, Y\leqslant\sqrt{t}\}\approx K_6F(t).$$

This gives the result

$$\Pr\{X\cdot Y\leqslant t\}\approx K_3t^{\gamma_F}\log(1/t),$$

which completes the proof of Proposition 2c. ∎

REFERENCES

Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1987). *Regular variation*. Cambridge, UK: Cambridge Univ. Press.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review,* **97**, 523–547.

Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology*, **34**, 253–275.

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

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton Univ. Press.

de Haan, L. (1990). Fighting the arch-enemy with mathematics. *Statistica Neerlandica*, **44**, 45–68.

Embrechts, P., & Goldie, C. M. (1980). On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society (Serie A)*, **29**, 243–256.

Feller, W. (1966). *An introduction to probability theory and its applications (Vol. II)*. New York: Wiley.

Fisher, R. A, & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics*. New York: Wiley.

Geluk, J. L. (1994). Asymptotic behaviour of the convolution tail of distributions each having a first or second order regularly varying tail. *Analysis*, **14**, 163–183.

Geluk, J. L. (1996). Tails of subordinated laws: The regularly varying case. *Stochastic Processes and Their Applications*, **61**, 147–161.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, **44**, 423–453.

Green, D. M., & Luce, R. D. (1975). Parallel psychometric functions from a set of independent detectors. *Psychological Review*, **82**, 483–486.

Gumbel, E. J. (1958). *The statistics of extremes*. New York: Columbia Univ. Press.

Indow, T. (1993). *Retention curves of artificial and natural memory: tight and soft models* (Tech. Rep. 93–11). Institute for Mathematical Behavioral Sciences, University of California, Irvine.

Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. New York: Springer-Verlag.

Link, S. W. (1992). Imitatio Estes: Stimulus sampling origin of Weber's law. In Healy, A. L., Kosslyn, S. M., & Shiffrin, R. M. (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (pp. 97–113). Hillsdale, NJ: Erlbaum.

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

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

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

Luce, R. D. (1986). *Response times, their role in inferring elementary mental organization*. New York: Oxford Univ. Press.

Marley, A. A. J. (1989). A random utility family that includes many of the "classical" models and has closed form choice probabilities and choice reaction times. *British Journal of Mathematical and Statistical Psychology*, **42**, 13–36.

Marley, A. A. J., & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretations. *Journal of Mathematical Psychology*, **36**, 1–20.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (Eds.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Shibuya, H., & Bundesen, C. (1988). Visual selection from multielements displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 591–600.

Ulrich, R., & Giray, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Perception and Psychophysics*, **39**, 248–254.

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review*, **2**, 20–54.

Wandell, B., & Luce, R. D. (1978). Pooling peripheral information: Averages versus extreme values. *Journal of Mathematical Psychology*, **17**, 220–235.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**, 292–297.

Wolfram, S. (1996). *The Mathematica book (3 ed.)*. New York: Cambridge Univ. Press.

Received: January 31, 2000; published online: January 10, 2002