Online Appendix

1 Motivation for the Type I Distribution

The interperson cdf G could conceivably belong to the domain of attraction of any of the three limiting distributions in extreme value theory. To determine which of these three distributions is most appropriate, we investigate whether the interperson scores are fat-tailed or not and whether they can be modeled as unbounded or not. To that end, we turn to the study in $\S$5.13 in [1], in which more than 4 million interperson similarity scores were computed by the National Institute for Standards and Technology (NIST). In Fig. 6 of the Online Supporting Information of [2], we plotted the interperson similarity score pdf generated by these data and fitted an exponential tail, a gamma and a lognormal distribution to the data. This figure shows that for the biometric matcher used in $\S$5.13 in [1], an exponential tail provides an excellent fit to the data, which precludes the possibility of G being in the domain of attraction of the Type II distribution. More specifically, G is well represented by

$$G(x) = 1 - ke^{-\lambda x} \quad \forall x \geq x^*, \tag{1}$$

where x^* is the point after which $G(x)$ exhibits exponential tail behavior. With $x^* = 6$, the maximum likelihood estimates of k and λ are 3.7042 and 0.3394, respectively. For these parameter values, we compute that for a gallery size of 6 billion (roughly the entire population of the world), the probability of the maximum of 6 billion iid interperson similarity scores exceeding a score of 115 is $< 10^{-6}$. In contrast, the biometric matcher used in study $\S$5.13 in [1] generates similarity scores on a scale of 0 to 499. Hence, for all practical purposes, G can be modeled as having an unbounded right tail, i.e., as being in the domain of attraction of the Type I distribution.

2 Derivation of Equation (9) in the Main Text

To derive equation (9) in the main text, it suffices to show that

$$\int_0^\infty \Big[a'_m + (\frac{x-a_m}{b_m})b'_m\Big]\frac{1}{b_m}\exp(-e^{-(x-a_m)/b_m})e^{-(x-a_m)/b_m}\sum_{i=0}^{r}\frac{f^{(i)}(a_m)(x-a_m)^i}{i!}\,dx$$
$$\approx a'_m\sum_{i=0}^{r}\frac{f^{(i)}(a_m)b_m^i\mu_i}{i!} + b'_m\sum_{i=0}^{r}\frac{f^{(i)}(a_m)b_m^i\mu_{i+1}}{i!}. \quad (2)$$

We first express the left side of (2) as

$$\int_0^\infty \Big[a'_m + (\frac{x-a_m}{b_m})b'_m\Big]\frac{1}{b_m}\exp(-e^{-(x-a_m)/b_m})e^{-(x-a_m)/b_m}\sum_{i=0}^{r}\frac{f^{(i)}(a_m)(x-a_m)^i}{i!}\,dx$$

$$= a'_m\sum_{i=0}^{r}\frac{f^{(i)}(a_m)}{i!}\int_0^\infty\frac{(x-a_m)^i}{b_m}\exp(-e^{-(x-a_m)/b_m})e^{-(x-a_m)/b_m}\,dx$$
$$+b'_m\sum_{i=0}^{r}\frac{f^{(i)}(a_m)}{i!}\int_0^\infty\frac{(x-a_m)^{i+1}}{b_m^2}\exp(-e^{-(x-a_m)/b_m})e^{-(x-a_m)/b_m}\,dx, \quad (3)$$
$$= a'_m\sum_{i=0}^{r}\frac{f^{(i)}(a_m)b_m^i}{i!}\int_0^\infty\left(\frac{x-a_m}{b_m}\right)^i\frac{1}{b_m}\exp(-e^{-(x-a_m)/b_m})e^{-(x-a_m)/b_m}\,dx$$
$$+b'_m\sum_{i=0}^{r}\frac{f^{(i)}(a_m)b_m^i}{i!}\int_0^\infty\left(\frac{x-a_m}{b_m}\right)^{i+1}\frac{1}{b_m}\exp(-e^{-(x-a_m)/b_m})e^{-(x-a_m)/b_m}\,dx. \quad (4)$$

Let $h(y) = \exp(-e^{-y})e^{-y}$ be the pdf of the standard Type I distribution with mode at 0 and scale parameter equal to 1 [3]. Using the change of variable $y = \frac{x-a_m}{b_m}$, we simplify the integrals in equation (4) to

$$\int_0^\infty\left(\frac{x-a_m}{b_m}\right)^i\frac{1}{b_m}\exp(-e^{-(x-a_m)/b_m})e^{-(x-a_m)/b_m}\,dx = \int_{-a_m/b_m}^\infty y^i\exp(-e^{-y})e^{-y}\,dy, \quad (5)$$
$$= \int_{-a_m/b_m}^\infty y^i h(y)\,dy, \quad (6)$$
$$\approx \int_{-\infty}^\infty y^i h(y)\,dy, \quad (7)$$
$$= \mu_i, \quad (8)$$

where μ_i is the i^{th} moment of the standard Type I distribution, and where the approximation in (7) is valid because $f(x)$ is diffuse in the domain where $h'_n(x)$ has most of its mass.

Using equation (8) to substitute for both integrals in equation (4) gives equation (2).

3 Derivation of Equation (12) in the Main Text

With the gamma intrapersonal pdf $f(x)$ from equation (11) in the main text, we have that $f^{(1)}(x) = \left[\frac{\alpha-1}{x} - \frac{1}{\beta}\right] f(x)$. With the exponential-tail interpersonal scores from equation (10) in the main text, we have that $a_n = \frac{\ln(kn)}{\lambda}$ and $b_n = \frac{1}{\lambda}$ (page 155 of [4]), and thus $a'_n = \frac{1}{n\lambda}$ and $b'_n = 0$. Substituting the mode a_n into the gamma pdf gives $f(a_n) = \frac{(kn)^{(-1/\lambda\beta)}\left(\ln(kn)\right)^{\alpha-1}}{\lambda^{\alpha-1}\beta^\alpha\Gamma(\alpha)}$. Making the above substitutions and setting $r = 1$ (and $\mu_0 = 1$, $\mu_1 = \gamma$) in equation (9) in the main text yield

$$\begin{aligned} d(n) &= \int_n^\infty \frac{1}{m\lambda}\left[1 - \frac{\gamma}{\lambda\beta} + \frac{\gamma(\alpha-1)}{\ln(km)}\right] \frac{(km)^{(-1/\lambda\beta)}\left(\ln(km)\right)^{\alpha-1}}{\lambda^{\alpha-1}\beta^\alpha\Gamma(\alpha)}\, dm, & (9) \\ &= \frac{\lambda\beta-\gamma}{(\lambda\beta)^{\alpha+1}\Gamma(\alpha)} \int_n^\infty \frac{\left(\ln(km)\right)^{\alpha-1}}{m(km)^{(1/\lambda\beta)}}\, dm + \frac{\gamma(\alpha-1)}{(\lambda\beta)^\alpha\Gamma(\alpha)} \int_n^\infty \frac{\left(\ln(km)\right)^{\alpha-2}}{m(km)^{(1/\lambda\beta)}}\, dm. & (10) \end{aligned}$$

To compute the integrals in equation (10), we introduce the upper incomplete gamma function $\Gamma(q, x)$ [5],

$$\Gamma(q, x) = \int_x^\infty t^{q-1} e^{-t}\, dt \quad \text{for} \quad x > 0. \tag{11}$$

A standard property of this function is that [5]

$$\Gamma(q+1, x) = q\Gamma(q, x) + x^q e^{-x}. \tag{12}$$

Now consider the change of variable $t = \frac{\ln(km)}{\lambda\beta}$ in the first integral in equation (10),

$$\begin{aligned} \int_n^\infty \frac{\left(\ln(km)\right)^{\alpha-1}}{m(km)^{(1/\lambda\beta)}}\, dm &= (\lambda\beta)^\alpha \int_{\frac{\ln(kn)}{\lambda\beta}}^\infty t^{\alpha-1} e^{-t}\, dt, & (13) \\ &= (\lambda\beta)^\alpha \Gamma(\alpha, \frac{\ln(kn)}{\lambda\beta}) \quad \text{by (11).} & (14) \end{aligned}$$

Substituting equation (14) for both integrals in (10) and recalling that $d(\infty) = 0$ yield

$$\begin{aligned} d(n) &= \frac{(\lambda\beta-\gamma)\Gamma(\alpha, \frac{\ln(kn)}{\lambda\beta})}{\lambda\beta\Gamma(\alpha)} + \frac{\gamma(\alpha-1)\Gamma(\alpha-1, \frac{\ln(kn)}{\lambda\beta})}{\lambda\beta\Gamma(\alpha)}, & (15) \\ &= \frac{(\lambda\beta-\gamma)}{\lambda\beta\Gamma(\alpha)}\left[(\alpha-1)\Gamma(\alpha-1, \frac{\ln(kn)}{\lambda\beta}) + \frac{(\ln(kn))^{\alpha-1}}{(\lambda\beta)^{\alpha-1}(kn)^{1/\lambda\beta}}\right] + \frac{\gamma(\alpha-1)\Gamma(\alpha-1, \frac{\ln(kn)}{\lambda\beta})}{\lambda\beta\Gamma(\alpha)} \quad \text{by (12),} & (16) \\ &= \frac{(\alpha-1)\Gamma(\alpha-1, \frac{\ln(kn)}{\lambda\beta})}{\Gamma(\alpha)} + \frac{(\lambda\beta-\gamma)\left(\ln(kn)\right)^{\alpha-1}}{(\lambda\beta)^\alpha\Gamma(\alpha)\left(kn\right)^{1/\lambda\beta}}. & (17) \end{aligned}$$

4 Derivation of Equation (13) in the Main Text

If $d(n)$ was truly log-linear in n, then $nd'(n)$ would be a negative constant. In contrast, for the distributional assumptions in equations (10) and (11) in the main text, $nd'(n)$ grows nearly linearly in $\ln(n)$ for $n \in [10^2, 10^9]$ (Fig. 3). Using Fig. 3, we approximate $nd'(n)$ by

$$nd'(n) \approx 1.68 \times 10^{-3} \ln(n) - 5.25 \times 10^{-2}, \tag{18}$$

and then integrate $d'(m)$ from $m = 10^2$ to $m = 10^9$ to get

$$\begin{aligned} d(n) \quad &\approx \quad d(100) + 0.2238 - 5.25 \times 10^{-2} \ln(n) + 8.4 \times 10^{-4} \big(\ln(n)\big)^2, &(19) \\ &\quad\ 0.7293 - 5.25 \times 10^{-2} \ln(n) + 8.4 \times 10^{-4} \big(\ln(n)\big)^2. &(20) \end{aligned}$$

5 Approximate Analysis of $\mathrm{d}_i(n)$

To use the generic results developed in equation (10) of the main text, we need an extreme value result for $\Omega_i(x)$, which is a mixture of Weibull distributions. Before doing so, we first state the extreme value result for the individual Weibull interperson distributions. On page 155 of [4], we find that $G_i(x)$ belongs to the domain of attraction of the Type I distribution with normalizing constants given by

$$a_{in} = \left(\frac{\ln(n)}{c_i}\right)^{(1/\tau_i)}, \tag{21}$$

$$b_{in} = \frac{1}{c_i \tau_i} \left(\frac{\ln(n)}{c_i}\right)^{(1/\tau_i - 1)}. \tag{22}$$

We now use a result derived in [6] to obtain the extreme value result for the mixture distribution $\Omega_i(x)$. Following [6], we first define tail dominance. The right tail of a distribution $\Phi_1(x)$ dominates the tail of another distribution $\Phi_2(x)$ if

$$\lim_{x \to \infty} \frac{1 - \Phi_2(x)}{1 - \Phi_1(x)} = \rho \tag{23}$$

for some $\rho \geq 0$. The distributions $\Phi_1(x)$ and $\Phi_2(x)$ are tail-equivalent if $\rho > 0$, and the tail of $\Phi_1(x)$ strictly dominates the tail of $\Phi_2(x)$ if $\rho = 0$ [6, 7].

The tail ratio of any two distributions in our interperson Weibull family is given by

$$\begin{aligned} \rho_{ij} &= \frac{1 - G_i(x)}{1 - G_j(x)}, && (24) \\ &= \frac{e^{-c_i x^{\tau_i}}}{e^{-c_j x^{\tau_j}}}, && (25) \\ &= e^{-x^{\tau_j}(c_i x^{\tau_i - \tau_j} - c_j)}, && (26) \end{aligned}$$

and hence for $\tau_i > \tau_j$,

$$\lim_{x \to \infty} \rho_{ij} = 0. \tag{27}$$

Thus $G_j(x)$ tail-dominates $G_i(x)$ if $\tau_j < \tau_i$. From Table 1 we see that τ_5 has the smallest numerical value followed by τ_8. By equation (18) in the main text, G_8 has the dominant tail in the mixture distributions Ω_6, Ω_7 and Ω_8, whereas G_5 has the dominant tail in the mixture distributions Ω_1, Ω_2, Ω_3, Ω_4 and Ω_5.

We present Theorem 2 of [6] in our notation. Let G^* be the dominant distribution in the mixture distribution Ω_i. Let G^* belong to the domain of attraction of the Type-I distribution with normalizing constants a_n^* and b_n^*. Then Ω_i also belongs to the domain of attraction of the Type-I distribution with normalizing constants given by

$$\begin{aligned} a_{in} &= a_n^* + b_n^* \log(\gamma_i), && (28) \\ b_{in} &= b_n^*, && (29) \end{aligned}$$

where

$$\gamma_i = \begin{cases} \sum_{j=1}^{i} p(j) & \text{if } G^* = G_i, \\ p(j) & \text{if } G^* = G_j\ ,\ \ j > i. \end{cases} \tag{30}$$

Substituting the relevant expressions from equations (21)-(22) into equations (28)-(29), we get the following expressions for the normalizing constants of the mixture distribution Ω_i:

$$a_{in} = \begin{cases} \left(\frac{\ln(n)}{c_8}\right)^{(1/\tau_8)} + \log(\gamma_i)\frac{1}{c_8\tau_8}\left(\frac{\ln(n)}{c_8}\right)^{(1/\tau_8-1)} & \text{if } i > 5, \\ \left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5)} + \log(\gamma_i)\frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)} & \text{if } i \leq 5, \end{cases} \tag{31}$$

$$b_{in} = \begin{cases} \frac{1}{c_8\tau_8}\left(\frac{\ln(n)}{c_8}\right)^{(1/\tau_8-1)} & \text{if } i > 5, \\ \frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)} & \text{if } i \leq 5. \end{cases} \tag{32}$$

We are now in a position to apply equation (9) in the main text. Recall that the approximations underlying equation (9) in the main text are accurate only when $f_i(x)$ is highly diffuse relative to $h_{in}(x)$ (see Fig. 1). Fig. 4.6 of [9] has 8 plots, one for each image quality, that are similar to Fig. 1, which indeed show that $f_i(x)$ is nearly linear where $h_{in}(x)$ has its mass for each image quality i. Hence, we can use the Taylor series expansion with confidence.

We now present expressions for a'_{1n}, b'_{1n}, and $f_1(a_{1n})$, which are needed in equation (9) in the main text. Differentiating a_{1n} and b_{1n} with respect to n in equations (31)-(32), we get

$$a'_{1n} = \frac{1}{n}\frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}\left(1 + \frac{(1-\tau_5)\ln(\gamma_1)}{\tau_5\ln(n)}\right), \tag{33}$$

$$b'_{1n} = \frac{1}{n}\frac{(1-\tau_5)}{(c_5\tau_5)^2}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-2)}. \tag{34}$$

Substituting the expression for a_{1n} into the right side of equation (14) in the main text yields

$$f_1(a_{1n}) = \frac{n^{-\left[\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)} + \ln(\gamma_1)\frac{1}{c_5\tau_5}\right]}}{\beta_1^{\alpha_1}\Gamma(\alpha_1)}\left[\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5)} + \log(\gamma_1)\frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}\right]^{\alpha_1-1}. \tag{35}$$

We again set $r = 1$ and employ a linear approximation of $f_1(x)$ in equation (9) in the main text. We use $f_1^{(1)}(x) = \left[\frac{\alpha_1-1}{x} - \frac{1}{\beta_1}\right]f_1(x)$ to evaluate the first derivative of $f_1(x)$ at a_{1n} and substitute the expressions for a'_{1n}, b'_{1n}, and $f_1(a_{1n})$ from above into equation (10) in the main text to get

$$
\begin{aligned}
d'_1(n) = -\Bigg[&\Bigg\{\frac{1}{n}\frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}\left(1+\frac{(1-\tau_5)\ln(\gamma_1)}{\tau_5\ln(n)}\right)\Bigg\}\times \\
&\Bigg\{1+\frac{\mu_1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}\Bigg[\frac{(\alpha_1-1)}{\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5)}+\ln(\gamma_i)\frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}}-\frac{1}{\beta_1}\Bigg]\Bigg\}\times \\
&\Bigg\{\frac{n^{-\left[\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}+\ln(\gamma_1)\frac{1}{c_5\tau_5}\right]}}{\beta_1^{\alpha_1}\Gamma(\alpha_1)}\Bigg[\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5)}+\ln(\gamma_1)\frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}\Bigg]^{\alpha_1-1}\Bigg\}\Bigg] \\
-\Bigg[&\Bigg\{\frac{1}{n}\frac{(1-\tau_5)}{(c_5\tau_5)^2}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-2)}\Bigg\}\times \\
&\Bigg\{\mu_1+\frac{\mu_2}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}\Bigg[\frac{(\alpha_1-1)}{\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5)}+\log(\gamma_1)\frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}}-\frac{1}{\beta_1}\Bigg]\Bigg\}\times \\
&\Bigg\{\frac{n^{-\left[\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}+\ln(\gamma_1)\frac{1}{c_5\tau_5}\right]}}{\beta_1^{\alpha_1}\Gamma(\alpha_1)}\Bigg[\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5)}+\log(\gamma_1)\frac{1}{c_5\tau_5}\left(\frac{\ln(n)}{c_5}\right)^{(1/\tau_5-1)}\Bigg]^{\alpha_1-1}\Bigg\}\Bigg].
\end{aligned}
\tag{36}
$$

We can develop expressions similar to equation (36) for other image qualities. The only change in the expressions would be the different values for the parameters of the intraperson distribution, the parameters for the dominant interperson distribution, and γ_i.

6 Parameter Estimation for the Example With Image Quality

In this section, we estimate the 16 parameters $(\alpha_i, \beta_i, \sigma_i, \theta_i)$, $i = 1, \ldots, 4$, in equations (14) and (22) in the main text, for each of the 18 matching systems in Table 1 of the main text. The raw data are the m_i points on the TAR vs. FAR quality-i curves in Figs. D20-D23 of [10]. For example, as seen in Fig. C-37 of [10], which has the 4 curves for the Cogent matcher from Figs. D20-D23, we use $m_1 = 4$, $m_2 = 5$, $m_3 = 13$, $m_4 = 16$.

Each TAR vs. FAR curve is generated by varying the threshold level. These threshold values are not reported in [10] and need to be estimated with the 16 parameters. For $j = 1, \ldots, m_i$, let t_{ij} be the unknown threshold corresponding to the j^{th} point on the quality-i TAR vs. FAR curve, and let the TAR and FAR for this point be denoted by d_{ij} and f_{ij}, respectively. According to our theoretical model, we have

$$1 - F_i(t_{ij}) = d_{ij}, \tag{37}$$

$$1 - \Big(\sum_{k=1}^{i} p(k)\Big)G_i(t_{ij}) - \sum_{k=i+1}^{4} p(k)G_k(t_{ij}) = f_{ij}, \tag{38}$$

where the right sides of (37)-(38) are raw data and the left sides of (37)-(38) contain unknown parameters and thresholds.

This estimation problem is ill-conditioned: because $P(\text{Gamma}(\alpha, \beta) < t) = P(\text{Gamma}(\alpha, k\beta) < kt)$ for some threshold t and for all $k > 0$, it follows that $(\alpha_i, \beta_i, \sigma_i, \theta_i, t_{ij})$ and $(\alpha_i, k\beta_i, \sigma_i, k\theta_i, kt_{ij})$ provide equally good (or bad) fits to the data for all $k > 0$. Hence, we need additional data to estimate one of these unknowns, and we use the quality-aggregated intraperson and interperson similarity scores in Figs. C-35 and C-36 of [10].

We proceed as follows. Because f_{ij} depends on $G_i, \ldots, G_4$ but not on $G_1, \ldots, G_{i-1}$ (see (38)), we solve for the parameters and threholds iteratively for each image quality, beginning with quality 4. We fit the quality-aggregated intraperson similarity scores in Fig. C-35 of [10] to a gamma distribution and set β_4 equal to the scale parameter resulting from this estimation. Using (37)-(38), we minimize the weighted normalized least squares quantity,

$$\sum_{j=1}^{m_4} w_{4j}(1 - F_4(t_{4j}) - d_{4j})^2 + \sum_{j=1}^{m_4} \tilde{w}_{4j}(1 - G_4(t_{4j}) - f_{4j})^2, \tag{39}$$

over $(\alpha_4, \sigma_4, \theta_4, t_{41}, \ldots, t_{4m_4})$, where the weights $w_{ij} = 10^{2\lceil -\log_{10}(d_{ij})+2\rceil}$ and $\tilde{w}_{ij} = 10^{2\lceil -\log_{10}(f_{ij})+2\rceil}$ are used to obtain sufficient accuracy for small TARs and FARs.

We then solve

$$\min_{\substack{\alpha_3,\beta_3,\sigma_3,\theta_3\\ t_{31},\dots,t_{3m_3}}} \sum_{j=1}^{m_3} w_{3j}(1 - F_3(t_{3j}) - d_{3j})^2 + \sum_{j=1}^{m_3} \tilde{w}_{3j} \left(1 - \big(\sum_{k=1}^{3} p(k)\big) G_3(t_{3j}) - p(4) G_4(t_{4j}) - f_{3j}\right)^2, \tag{40}$$

followed by

$$\min_{\substack{\alpha_2,\beta_2,\sigma_2,\theta_2\\ t_{21},\dots,t_{2m_2}}} \sum_{j=1}^{m_2} w_{2j}\left(1 - F_2(t_{2j}) - d_{2j}\right)^2 + \sum_{j=1}^{m_2} \tilde{w}_{2j} \left(1 - \big(\sum_{k=1}^{2} p(k)\big) G_2(t_{2j}) - \sum_{k=3}^{4} p(k) G_k(t_{kj}) - f_{2j}\right)^2, \tag{41}$$

and finally

$$\min_{\substack{\alpha_1,\beta_1,\sigma_1,\theta_1\\ t_{11},\dots,t_{1m_1}}} \sum_{j=1}^{m_1} w_{1j}\left(1 - F_1(t_{1j}) - d_{1j}\right)^2 + \sum_{j=1}^{m_1} \tilde{w}_{1j} \left(1 - p(1) G_1(t_{1j}) - \sum_{k=2}^{4} p(k) G_k(t_{kj}) - f_{1j}\right)^2. \tag{42}$$

Finally, we search for a k such that the parameters $(\alpha_i, k\beta_i, \sigma_i, k\theta_i, kt_{ij})$ provide a best fit to the quality-aggregated intraperson and interperson similarity scores. That is, we choose k that maximizes the likelihood function of the data given in Figs. C-35 and C-36 of [10], with equal weight given to intraperson and interperson scores.

The 8 intraperson parameter values (α_i, β_i), $i = 1, \dots, 4$, and the 8 interperson parameter values (σ_i, θ_i), $i = 1, \dots, 4$, for each of the 18 matching systems appear in Tables 2 and 3, respectively.

References

[1] Wilson C. L., Watson C. I., Garris M. D. & Hicklin A. (2003) *Studies of fingerprint matching using the NIST verification test bed (VTB), NISTIR 7020* (National Institute of Standards and Technology, Gaithersburg, MD).

[2] Wein L. M., Baveja M. (2005) Using fingerprint image quality to improve the identification performance of the U.S. Visitor and Immigrant Status Indicator Technology Program. *PNAS* **102**, 7772-7775.

[3] Johnson N. L., Kotz S., Balakrishnan N. (1995) *Continuous Univariate Distributions*, V.1-2, John-Wiley, New York.

[4] Embrechts P. (1997) *Modelling Extremal Events*, Springer, Berlin.

[5] Abramowitz M., Stegun I. A. (1964) *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*, Dover.

[6] Kang S., Serfozo R. F. (1999) Extreme Values of Phase-Type and Mixed Random Variables with Parallel-Processing Examples. *J. Appl. Prob.* **36**. 194-210.

[7] Resnick S. I. (1971) Tail Equivalence and Its Applications. *J. Appl. Prob.* **8**, 136-156.

[8] Galambos, J. (1987) *The Asymptotic Theory of Extreme Order Statistics*, Krieger.

[9] Baveja, M. (2007) An Analysis of the US-VISIT Biometric Program. Ph.D. thesis, Institute of Computational and Mathematical Engineering, Stanford University, Stanford, CA.

[10] Wilson C. L., Hicklin R. A., Korves H., Ulery B., Zoepfl M., Bone M., Grother P. Michaels R., Otto S. & Watson C. (2004) *Fingerprint Vendor Technology Evaluation 2003: Summary of Results and Analysis Report* (National Institute of Standards and Technology, Gaithersburg, MD), National Institute of Standards and Technology Internal Report 7123.

Quality i	c_i	τ_i
1	2.3973×10^{-6}	2.2564
2	2.9760×10^{-3}	1.2458
3	3.8013×10^{-2}	0.8892
4	6.3144×10^{-3}	1.1374
5	1.1095×10^{0}	0.4183
6	4.5238×10^{-3}	1.1806
7	3.7081×10^{-2}	0.8862
8	7.3230×10^{-1}	0.4713

Table 1 Parameter values for the eight Weibull interperson similarity score distributions.

Matcher	α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4
NEC	3.72	2.00	2.25	0.32	2267.2	22698.6	3616.2	24490.5
Cogent	6.53	8.76	1.65	0.98	506.0	281.6	4145.2	1482.3
SAGEM M2	4.91	5.32	6.61	1.54	4244.8	3473.6	1471.4	1902.4
SAGEM M1	3.10	4.71	9.96	5.78	7981.6	2346.4	508.4	811.0
Neurotech M1	1.46	2.35	1.24	1.06	275.6	55.6	77.2	49.3
Motorola	3.61	8.61	12.1	6.06	46.9	9.77	4.90	8.57
Identix	2.12	3.54	3.18	0.12	256.3	58.1	44.6	1067.5
NIST VTB	2.51	3.17	0.87	0.18	35.1	17.9	41.0	25.6
UltraScan M1	1.44	2.04	0.83	0.49	2.86	0.87	3.07	2.67
UltraScan M2	0.68	0.87	0.48	0.40	99.8	16.9	26.6	23.7
Biolink	2.78	3.06	2.10	0.08	351.4	188.0	149.9	5375.2
Antheus	1.03	1.27	0.10	0.09	327089.9	106444.3	45616.0	24528.9
Phoenix	1.79	0.60	0.31	0.50	325937.0	1326859.2	743610.6	105468.8
Technoimagia	2.69	42.5	8.19	1.47	0.23	0.0047	0.017	0.076
123 ID M2	1.30	1.25	0.71	0.61	68181215.9	42878691.1	45300578.9	40508473.5
Golden Finger	1.34	1.47	1.48	0.54	99.4	39.7	34.7	17.4
Raytheon	1.52	1.01	1.50	0.51	0.70	0.69	0.32	0.07
Avalon	4.03	2.63	2.59	2.83	547.0	655.1	464.0	329.4

Table 2 The intraperson parameter values for the 18 fingerprint matching systems.

Matcher	σ_1	σ_2	σ_3	σ_4	θ_1	θ_2	θ_3	θ_4
NEC	1.58	0.54	0.48	8.74E-4	93.7	86.9	108.0	69.8
Cogent	0.42	3.56	0.37	0.40	18.7	59.2	100.3	63.2
SAGEM M2	7.07	8.96	11.6	7.59	240.1	198.8	153.1	56.3
SAGEM M1	12.3	8.88	11.1	57.8	128.5	130.3	26.9	38.7
Neurotech M1	2.93	2.46	4.70	4.61	1.75	2.41	1.27	2.61
Motorola	49.0	163.8	176.0	196.9	0.58	0.21	0.20	0.18
Identix	8.72	15.0	19.3	0.04	4.56	2.86	2.25	1.19
NIST VTB	11.4	6.56	0.98	0.08	0.74	1.51	1.58	0.15
UltraScan M1	1.91	3.57	0.55	0.97	0.04	0.04	0.12	0.12
UltraScan M2	0.34	0.38	0.22	0.30	0.10	0.22	0.27	0.87
Biolink	11.8	12.0	12.8	0.11	10.5	9.88	8.66	128.8
Antheus	4.09	2.28	0.06	0.08	2289.3	5038.2	1.56	111.9
Phoenix	21.3	2.52	0.57	6.01	4528.9	18563.2	14786.5	3198.9
Technoimagia	30.5	1009.4	119.8	5.49	0.00461	0.000171	0.00108	0.0195
123 ID M2	7.57	4.94	3.00	6.25	913528.9	1476223.3	1725659.3	2204127.6
Golden Finger	1.31	2.94	1.67	1.04	2.66	1.96	5.42	0.98
Raytheon	3.43	2.38	2.25	0.94	0.02	0.01	0.04	0.004
Avalon	18.9	7.49	8.00	7.17	37.0	66.1	86.8	102.2

Table 3 The interperson parameter values for the 18 fingerprint matching systems.

Figure Legends

Fig. 1. Plots of the interperson pdf $g(x)$ from equation (10) in the main text, the intraperson pdf $f(x)$ from equation (11) in the main text, and the pdf $h_n(x)$ of the maximum interperson score in the gallery for $n = 10^6$. For large n, the pdf $f(x)$ is nearly linear in the domain where $h_n(x)$ has most of its mass, motivating our low-degree Taylor series expansion of $f(x)$ about the mode of $h_n(x)$.

Fig. 2. The rank-one identification probability $d(n)$ vs. the gallery size n under the distributional assumptions in equations (10)-(11) in the main text. The exact $d(n)$ in equation (1) of the main text and the approximation in equation (12) in the main text are indistinguishable in this plot.

Fig. 3. The relationship between $nd'(n)$ and n for gallery sizes between 10^2 and 10^9.

Fig. 4. The relationship between $d_i(n)$ and gallery size n for quality image $i = 1, \ldots, 8$ in plots **(a)**,. . .,**(h)**.

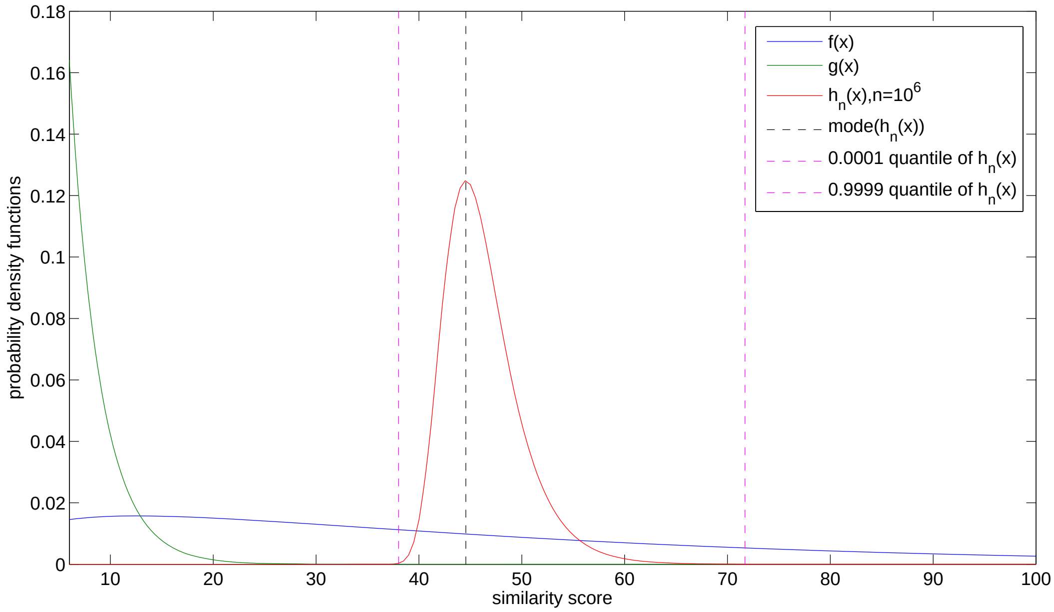

Figure 1

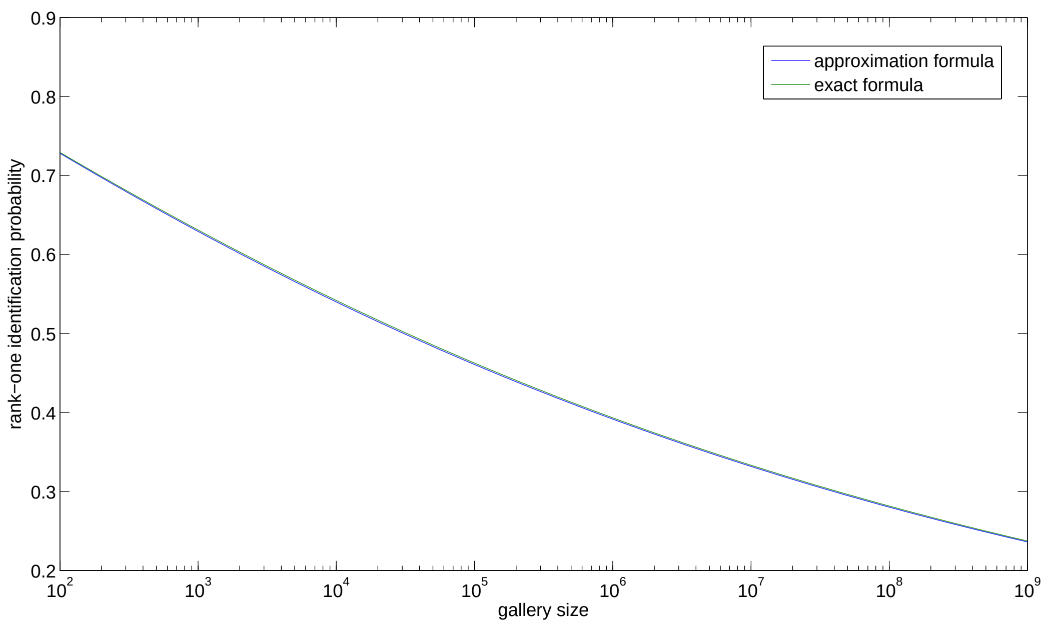

Figure 2

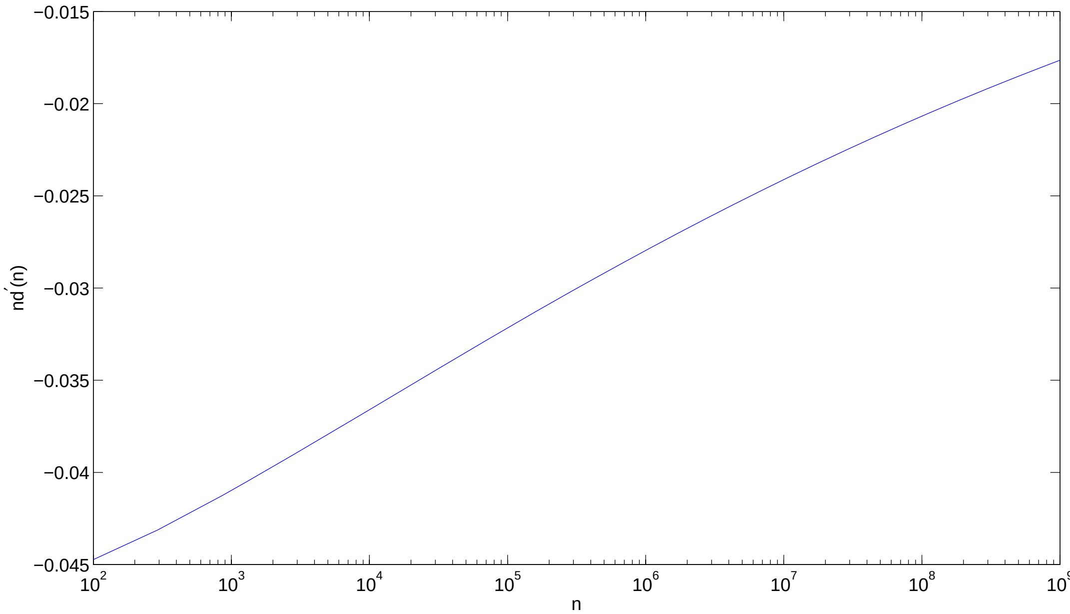

Figure 3

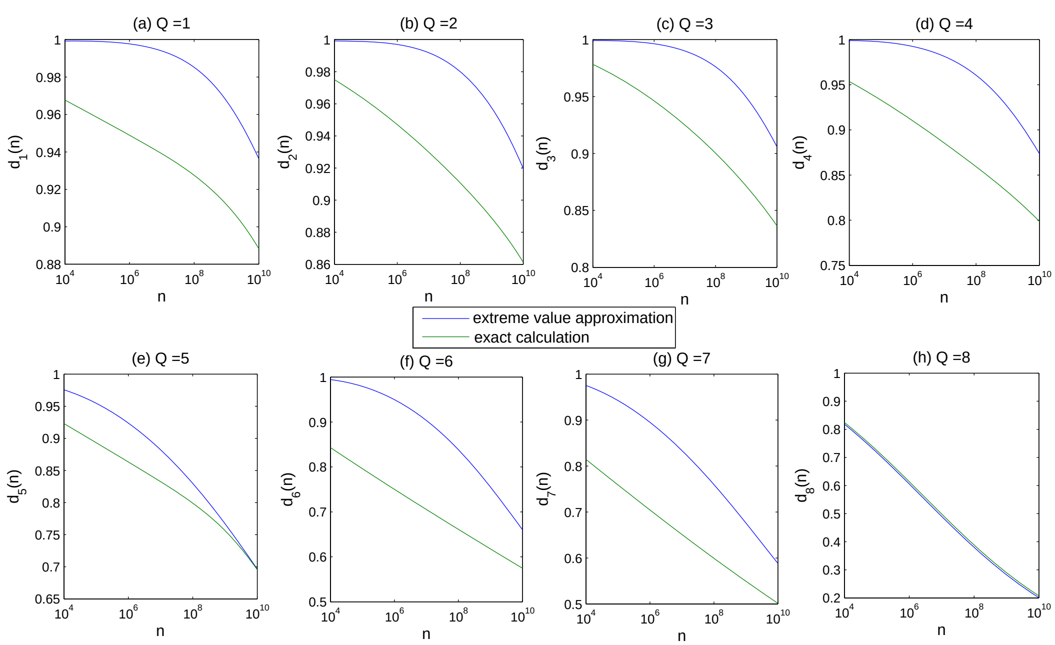

Figure 4