DOI: 10.36227/TECHRXIV.14061671.V1

Collision Avoidance on National and Global Scales: Understanding and Using Big Biometric Entropy

John Daugman[1]

***Abstract*—This short note discusses a biometric analog of the "birthday problem", when identity collisions within a group of random persons become likelier than not as group size grows. The general solution to this combinatorial problem is derived as a function of the verification False Match Rate. Its application to unique biometric identification on the planetary scale of the global human population is shown, referencing empirical data from US NIST (National Institute of Standards and Technology) trials involving 1.2 trillion (1.2×10^{12}) biometric comparisons. The entropy of both iris patterns suffices for global uniqueness.**

I. Introduction

APPLICANTS for Cambridge University undergraduate studies in mathematics or computer science are asked sometimes in their College interviews to reason about the "birthday problem": how many people, chosen at random, must be assembled until it becomes more likely than not that at least one pair of them have the same birthday? Some students are surprised that the answer is only 23 people. Although arriving at the exact number requires a calculator, the reasoning is that N people make $N(N-1)/2$ possible pairings. Given that each pairing has probability $1/365$ of sharing their birthday and $364/365$ of not, the probability that *none* of the pairings share a birthday is approximately $(364/365)^{N(N-1)/2}$, which is < 0.5 once $N \geq 23$.

There is a clear analogy with biometric collision avoidance, which we can formulate as the:

> **Biometric birthday problem**: if some biometric technology is operating with a verification FMR ("one-to-one" False Match Rate), how many people, chosen at random, must be assembled until it becomes more likely than not that at least one pair of them have a biometric collision (are falsely matched to each other)?

A good example is face recognition, tested across a wide variety of scenarios and using a broad range of image quality, but for which a high level of performance corresponds to making just one verification False Match in 1,000 non-mated comparisons [1] [2] [3]. That accuracy standard is better than human (even "super-recogniser") performance in some circumstances [3]. Face recognition algorithms have improved greatly in recent years, in terms of Rank-1 identification rates [1] [2] in test protocols in which a correct match does always exist within a search gallery that is populated also with other "distractors". But even in the recent tests, the best algorithms do still make some False Matches to distractor images even when there are only 100 distractors [1] [2] despite the presence of a correct match within the gallery, that should instead actually be returned at Rank-1.

Let us now consider the "biometric birthday problem" for a face recognition algorithm performing at FMR = 0.001 when examining a gallery of non-mated faces. How large must this gallery get before False Matches become likelier than not, in all-against-all comparisons? The answer is just 38. That number creates $38 \cdot 37/2 = 703$ possible pairings to consider, and $(1-0.001)^{703} = 0.495$ so False Matches are then already likelier than not. When waiting at Passport Control (or some other such queue), it is entertaining to turn around, look at the first 38 persons standing behind oneself, and try to spot the pair of facial doppelgängers [4] among them.

Biometric deployments at a national or even prospectively at the planetary scale face a massively challenging biometric "birthday problem" if they need to search for any duplicate identities, as was necessary in India when all 1.3 billion citizens were recently enrolled in a national ID programme for welfare distribution, government services, and subsidies (UIDAI: Unique IDentification Authority of India) [5]. Because enrollees had an incentive to acquire multiple identities and thereby issuance of multiple subsidies, every new enrollment had to be compared against all existing enrollments before an Aadhaar would be issued. This amounts to a search all-against-all for identity collisions among an astronomical $N(N-1)/2$ pairings of persons. Obviously any attempt to do this by face recognition would drown in False Matches from the very beginning. There simply is not enough entropy, or randomness, in human face structure; the necessary functional purposes of major facial features (mouth, nose, ocular areas) constrain their possible randomness. The bilateral symmetry normally present in a face further reduces its entropy by half. The key idea, the fundamental factor underlying the power of biometric identification, is entropy [6] [7].

Weak biometrics may be sufficient to enable "one-to-one" verification; stronger biometrics may enable identification in a search database of size N, "one-to-few" or "one-to-many" depending on N; but de-duplication applications exemplify the birthday problem in that they are essentially "all-against-all", and the number of False Match opportunities they must survive grows massively with N. In such deployments on a national scale, falsely detected or undetected identity collisions (even if few in percentage) would lead to reduced public confidence in and acceptance of the system, its impaired functionality, and legal problems caused both by undetected duplicates and falsely detected ones. Table I presents, for a broad range of FMR levels spanning 15 orders-of-magnitude, how large N

[1] Department of Computer Science and Technology, Cambridge University, Cambridge, UNITED KINGDOM. e-mail: John.Daugman@CL.cam.ac.uk

TABLE I
ACCURACY REQUIREMENTS FOR BIOMETRIC COLLISION AVOIDANCE

Verification FMR	Critical Population Size N
0.001	38 persons
0.0001	119 persons
10^{-5}	373 persons
10^{-6}	1,177 persons
10^{-9}	37,229 persons
10^{-12}	1.2 million persons
10^{-15}	37 million persons
10^{-18}	1.2 billion persons

can get before collisions become likelier than not. Table I clearly shows that the demands for a minuscule FMR become extremely daunting once the population size N is even that of a small town, let alone a population of national, continental, or of planetary scale.

II. General Solution for Population Bounds

The number of pairings possible among N persons is $N(N-1)/2$ because each person can be paired with $N-1$ others, but half of these are redundant (e.g. Alice and Bob, then also Bob and Alice); hence the halving. If a biometric technology is operating at some verification False Match Rate FMR, then the probability of a given pairing *not* resulting in a False Match is $(1-\mathrm{FMR})$, and the probability that *none* of the possible pairings do so is $(1-\mathrm{FMR})^{N(N-1)/2}$. For what value of N does this expression become < 0.5, and therefore a biometric collision becomes likelier than not?

We will invoke a property of the base e "natural logarithm" function $\log_{e=2.718...}(\ )$, commonly denoted $\ln(\ )$. We seek:

$$(1-\mathrm{FMR})^{N(N-1)/2} \quad < \quad 0.5 \tag{1}$$

$$\ln\left((1-\mathrm{FMR})^{N(N-1)/2}\right) \quad < \quad \ln(0.5) \tag{2}$$

$$\frac{N(N-1)}{2}\ln(1-\mathrm{FMR}) \quad < \quad -0.693 \tag{3}$$

Now using the power series expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots\ , \tag{4}$$

we have $\ln(1+x) \approx x$ for small $|x|$, whether $x \geq 0$ or $x < 0$. Basically this reflects the fact that the logarithm function is linear near where it crosses 0 at $\log(1)$, and the slope of this line is 1 if the base of the logarithm is e. Thus for any small FMR (say < 0.01), which also entails that $N^2 \gg N$, we have

$$-\frac{N(N-1)}{2}\ \mathrm{FMR} \quad \lesssim \quad -0.693 \tag{5}$$

$$N^2 \quad \gtrsim \quad 1.386/\mathrm{FMR} \tag{6}$$

$$N \quad \gtrsim \quad \sqrt{1.386/\mathrm{FMR}} \tag{7}$$

This general (but approximated) solution can be confirmed by evaluating (1) exactly, using for N each of the corresponding FMR cases tabulated in Table I, insofar as the available tools of calculation can handle the combinatorial exponents required in (1) when N is large.

III. Biometric Entropy to the Rescue

Entropy measures the complexity and randomness [6] that is present in (and between) random variables. Facial structure has limited capacity for randomness. The major facial features have a canonical standard configuration, usually with bilateral symmetry; the eyes are normally on opposite sides of the nose. Much greater randomness is found in iris patterns, and this is the origin of their legendary resistance to False Matches. Although often there do exist strong radial correlations within an iris, with mutual information as large as 0.3 bits per bit across radius [8], and also IrisCode bits at adjacent or nearby angles but a shared radial coordinate have "sticky oscillator" correlations that reduce their entropy as much as 0.5 bits per bit [7], nevertheless the remaining entropy is vast. Fig. 1 illustrates this graphically in the bit streams that constitute the IrisCodes of four different eyes. How IrisCodes are computed has been revealed previously [10]. The two bit values are equiprobable, so when bits in IrisCodes from two different eyes are compared by XOR (Exclusive-OR) to detect whether they agree or disagree, these outcomes again are equiprobable, amounting to the toss of a fair coin.

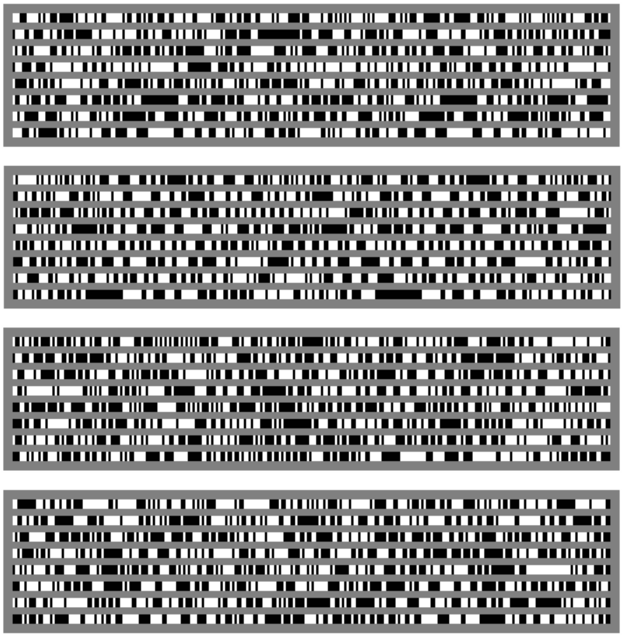

Fig. 1. Representation of the IrisCodes [10] produced by four different eyes. The eight rows within each can be regarded as eight concentric rings, each encoding a $[0, 2\pi]$ traversal around the iris. (Eyelid masking is not shown.)

The non-independence among the bits in a given IrisCode reduces their collective entropy from what would have been a maximum of 2,048 bits (if each bit corresponded to an independent "fair coin toss" Bernoulli trial) to only about 245 bits. Modelled as a "sticky oscillator" Markov process [7], IrisCode bits exhibit a phase coherence that can persist across several bits. Despite such losses in entropy, enough entropy remains that the collision probability between two IrisCodes from different eyes attenuates by astronomical factors, for small reductions in the tolerated accidental agreement.

IV. Discussion

A good way to understand this effect intuitively is to consider tossing a fair coin in runs of 245 tosses, tallying each run's fraction of heads. The total number of possible outcome sequences is 2^{245} and each of these has the same probability, namely $p_i = 2^{-245}$ (including, say, the "all heads" sequence). The entropy [6] contained in these possible sequences is:

$$\begin{aligned} H &= -\sum_i p_i \log_2(p_i) \qquad (8) \\ &= -\sum_{i=1}^{2^{245}} 2^{-245} \log_2(2^{-245}) = 245 \text{ bits.} \qquad (9) \end{aligned}$$

The vast majority of these sequences will have a nearly equal mix of heads and tails. The fraction of possible sequences that have (say) fewer than 30% heads is less than one-billionth of the total. This combinatorial property when large entropy (245 bits) exists in a random variable is ultimately the reason why, for iris recognition, a match between two IrisCodes can be accepted even when (say) 30% of their bits disagree due to problematic image acquisition. Despite such a lenient criterion being so tolerant of noisy bits, the probability that such an accepted match would actually be a False Match is, indeed, less than 1 in a billion.

The huge exponents appearing in (9) (note that $2^{245} \approx 10^{74}$) are key to understanding why sufficient entropy is the basis for biometric collision avoidance even at a planetary scale. A detailed tabulation of the relevant probability distributions, both densities and their cumulatives, with and without selecting for best matches after multiple image rotations to compensate for unknown head and camera tilt, is provided at [9] as a function of Hamming distance HD (fraction of bits that disagree in IrisCodes from two different eyes). This probability table enables us to predict how tolerant we can be of poor image acquisition (how large a fraction HD of disagreeing bits we can tolerate and still declare a match), without resulting in False Matches. The table [9] shows for acceptance criteria HD the resulting False Match probability, and its $\log_{10}$ (last two columns).

Table II extracts coarser HD increments of 0.01 from [9] (first column), showing the corresponding FMR predictions (second column). By 2003 image databases were only large enough to perform about 10 million iris cross-comparisons [10] but distribution parameters could be estimated, implying 249 bits of entropy (slightly more than 245), predicting FMR performance very similar to what is shown in Table II. No False Matches were observed below roughly the HD = 0.33 criterion, given the small databases available. The predicted FMR values were dismissed with incredulity by P.J. Phillips *inter alia* [11], because such FMR performance was unknown in other biometrics. But subsequently, other NIST researchers did actually perform billions [12] and then more than a trillion iris comparisons [13], obtaining FMR values in good agreement with those predictions, as reported in column 3.

TABLE II
False Match Rates Predicted in [9], and as Measured by NIST [12] with 1.16 Billion Iris Comparisons, and [13] with 1.2 Trillion Iris Comparisons

HD criterion	*FMR predicted in* [9]	*NIST* [12] [13] *measured FMR*
0.36	1 in 24,000	1 in 25,000
0.35	1 in 110,000	1 in 71,000
0.34	1 in 556,000	1 in 476,000
0.33	1 in 3.1 million	1 in 3.4 million
0.32	1 in 20 million	1 in 24 million
0.31	1 in 137 million	1 in 165 million
0.30	1 in 1.1 billion	1 in 2 billion
0.29	1 in 9 billion	(not measured)
0.28	1 in 92 billion	1 in 40 billion

An important cause of skepticism about the FMR performance levels shown in Table II, before they were eventually confirmed by NIST, was the existence of 'ground-truth' errors in early biometric databases that had created illusory identity collisions. Reports had emerged from the National Biometric Test Center (NBTC) in San Jose, California, saying that many False Matches occurred with iris recognition. This author requested and received the NBTC iris image database, which did indeed generate many apparent False Matches; but closer inspection revealed that they all were actually *true* matches, with ground-truth identity errors. After study, the Director of the NBTC admitted this was so, explaining: *"Clearly we were getting scammed by some of our student volunteers; they were changing names and coming through multiple times."* Apart from such extraordinarily sloppy and naïve data collection, which set back for years an appreciation of the discriminating powers of iris recognition, there is an inherent risk in estimating FMR by *intra*-dataset cross-comparisons. If even just one of N subjects is enrolled under two different identities, whether deviously or just through an innocent clerical error, the estimated FMR then cannot be better than $2/N^2$. The measured threshold calibration of FMR such as tabulated in Table II must then approach a floor, corresponding to this illusory FMR, which cannot be reduced by any reasonable change in threshold, and indeed NIST [12] demonstrated this problem for (university-sourced) *intra*-dataset comparisons.

NIST overcame this problem by performing *inter*-dataset comparisons: if two disjoint populations, of sizes (say) N and M in geographically remote places can be biometrically enrolled, then $N \times M$ inter-comparisons become possible without the contaminating effect of ground-truth errors. NIST [13] acquired enrollment datasets for two populations "very well separated geographically and occupationally," one having 3.9 million iris images used as the gallery, and the other having 315,000 iris images used as probes to search against this entire gallery, asserting there was zero likelihood of co-membership. Thereby NIST performed $N \times M = 1.2$ trillion IrisCode comparisons, leading to the FMR results shown in Table II column 3 (from [13] p. 61) for various HD threshold criteria. This close confirmation of theory (column 2), manipulating FMR over more than a million-fold range, is striking.

V. Conclusion

Iris recognition is perhaps unique among biometrics in having clear mathematical foundations, enabling strong predictions about IrisCode collision likelihood as a function of the decision threshold. As shown in Table II, for decision criteria in which no more than about 31% of the IrisCode bits are allowed to disagree when declaring a match (which is a very noise-tolerant criterion), the predicted FMR attenuates by about a factor of 10 for each additional 1% reduction in the tolerated amount of bit disagreement. This extraordinary fact seems not to be generally understood or appreciated; but it is a direct result of using high-entropy random variables in biometric codes, and of the combinatorial analysis of bit strings as summarised earlier. A critical lesson emerging here is the same as a lesson from cryptography: the great power of randomness, if you can get enough of it.

As confirmed independently by NIST in [13], the slope of the IrisCode Decision Error Trade-off curves is so flat that the FMR can be lowered by a factor of 10,000 to 100,000 while not even doubling the False non-Match rate (FnMR). A consequence of this relationship is that only small costs in increased FnMR need be paid, by HD threshold lowering, in order to increase greatly the size of a biometrically enrolled population without suffering collisions. Thus for IrisCodes from any two different eyes, the probability of HD ≤ 0.29 is about 10^{-10}. If we also exploit the fact that a person's two eyes generate IrisCodes that are almost completely independent, specifying 0.29 as a match criterion *binocularly* would yield a fusion FMR of about 10^{-20}. Equation (7) shows us that this is how the planetary human population can survive the "biometric birthday problem": it is unlikely that even a single pairing among 12 billion persons (despite the vast numbers of possible pairings) would disagree in $\leq 29\%$ of their IrisCode bits for both pairs of eyes. Thus speaks biometric entropy.

References

[1] I. Kemelmacher-Shlizerman, S.M. Seitz, D. Miller, and E. Brossard, "The MegaFace Benchmark: 1 million faces for recognition at scale," *Int'l. Conf. Comp. Vision & Patt. Recog.*, pp. 4873–4882, 2016.

[2] P. Grother, M. Ngan, and K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT), Part 2: Identification", NISTIR 8238, National Institute of Standards and Technology (Bethesda), 2018.

[3] P.J. Phillips, A. Yates, Y. Hu, C. Hahn, E. Noyes, K. Jackson, J. Cavazos, G. Jeckeln, R. Ranjan, S. Sankaranarayanan, J. Chen, C. Castillo, R. Chellappa, D. White, and A.J. O'Toole, "Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms," *Proc. Nat'l. Acad. Sci.*, vol. 115 (24), pp. 6171–6176, 2018.

[4] http://www.CL.cam.ac.uk/users/jgd1000/Doppelganger-photos.pdf

[5] S. Aiyar, *AADHAAR: A Biometric History of India's 12-Digit Revolution*. New Delhi: Westland Publications, 2017.

[6] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley-Interscience, 2006.

[7] J. Daugman, "Information Theory and the IrisCode," *IEEE Trans. Info. For. Sec*, vol. 11 (2), pp. 400–409, 2015.

[8] J. Daugman and C. Downing, "Radial correlations in iris patterns, and mutual information within IrisCodes", *IET Biometrics*, vol. 8 (3), pp. 185–189, 2019.

[9] http://www.CL.cam.ac.uk/users/jgd1000/IrisCumulatives.pdf

[10] J. Daugman, "The importance of being random: statistical principles of iris recognition," *Pattern Recognition*, vol. 36, pp. 279–291, 2003.

[11] P.J. Phillips, *inter alia*, comments at conferences and in publications, to UK Government Senior Biometrics Advisors, and to licensees of these iris recognition algorithms, as quoted to this author.

[12] P. Grother, E. Tabassi, G.W. Quinn, and W. Salamon, "IREX-I: Performance of Iris Recognition Algorithms on Standard Images." *NIST Interagency Report 7629*, 2009. Data is taken from Fig. 14 (page 46) and Table 7 (page 48) for 1.16 Billion inter-dataset iris comparisons.

[13] P. Grother, G.W. Quinn, J.R. Matey, M. Ngan, W. Salamon, G. Fiumara, and C. Watson, "IREX-III: Performance of Iris Identification Algorithms." *NIST Interagency Report 7836*, April 6, 2012.

John Daugman received his degrees at Harvard University and then taught at Harvard before coming to Cambridge University, where he is Professor of Computer Vision and Pattern Recognition. He has held the Johann Bernoulli Chair of Mathematics and Informatics at the University of Groningen, and the Toshiba Endowed Chair at the Tokyo Institute of Technology. His areas of research and teaching at Cambridge include computer vision, information theory, neural computing, and statistical pattern recognition. Awards for his work in science and technology include the Information Technology Award and Medal of the British Computer Society, the "Time 100" Innovators Award, and the OBE, Order of the British Empire. He has been elected a Fellow of: the Royal Academy of Engineering; the US National Academy of Inventors; the Institute of Mathematics and its Applications; the British Computer Society; and he has been inducted into the US National Inventors Hall of Fame. He is the founder and benefactor of the Cambridge Chrysalis Trust. Here he is represented by a sparse sum of 2D Gabor wavelets in six orientations and five frequencies.