Writing Technology-neutral Law
An Instructive Example

HANS VERGHESE MATHEWS

The regulation of "biotechnology" seems to present legislative difficulties of another kind. The Government of India proposes to institute a biotechnology regulatory authority—the Human DNA Profiling Bill 2015—for the purpose of collating the DNA information of citizens so as to strengthen the criminal justice process and system. However, the regulation of such technology needs to be technologically neutral and requires safeguards against any information abuse and exploitation on the part of the "experts" and "specialists" appointed to the new regulatory institution governing biotechnology.

Hans Verghese Mathews (*hans@cis-india.org*) is with the Centre for Internet and Society, Bengaluru.

Legislation that seeks to regulate technology must be up for scrutiny by judges and lawyers. And since the inner workings of a technology will be familiar only to specialists and expert practitioners, the phrasing of such laws must avoid technical details where possible. When the uses and products of technology can be specified closely, one feels such details can well be ignored and the requirement that the writing of law should be "technology-neutral," in this way, seems innocuous enough. But the uses and products of technology will seldom be full of technical details; with an emergent or a changeable technology the neutrality required of the law might concede too much discretionary power to the specialists. But it is in such circumstances, precisely, that neutrality is a desideratum: for technical details would necessitate frequent revisions in law that seeks to regulate a mutating "techne"—the frequent revision of a law would not be a good thing.

To consider the matter generally would be difficult, and we shall only look at a particular example in this article. The Government of India is proposing a Human DNA (Deoxyribose Nucleic Acid) Profiling Act, the preamble to which declares that its coordinate purposes are to, one, regulate the use of analysis of human body substance profiles for human identification; two, establish a DNA Profiling Board that would lay down appropriate standards for laboratories in terms of collecting human body substances and managing the custody trail from collection to reporting; and three, establish a National DNA Data Bank.

The phrasing of the first purpose is tortuous. Let us try to disentangle the words: Tissues and fluids are the generic "substances" of the human body, and "Deoxyribose Nucleic Acid" or DNA can be extracted from them—that is, from the cells typically found in tissues or fluids. The "analysis" of DNA serves many purposes. To serve as an identifier of the human person from whose body some quantum of DNA is extracted, is one such purpose; and the artefacts produced by such analysis have come to be called "DNA profiles." A person whose body yields "substance" for a DNA profile might be called the "organic source" of that profile. It must be noted that the organic source of a given profile may or may not be otherwise known to those who have the profile in their keeping. Why the word "artefact" has been resorted to, should become clear when we consider the DNA profile that have come into standard use—symbolic artefacts of a distinctive sort, we shall see.

The extraction of DNA and the assembly of DNA profiles are complex laboratory processes, and that a "DNA Profiling Board" should oversee them, seems only prudent. The securing of the "custody trail"—that the second purpose of this purported policy speaks of—is necessitated by the complex process of extraction and assembly without which the use of DNA profiles, for the purposes of identification in a court of law particularly, would be defeated. When they are symbolic artefacts, DNA profiles could be regarded as "data" deriving from DNA while as material artefacts obtained from the analysis of DNA, they could serve as identification profiles. Also, the circumstance that the third purpose of the policy mandates is the establishment of a data bank—rather than a "National DNA Profile Bank"—might be due to caution only.

In an annexure to the bill, which articulates the proposed legislation, the bill's proponents detail the process for the assembly and employment of volunteered DNA profiles—an undertaking which the epithet "national" seems to authorise, and the purpose of which we shall consider in good time. These volunteered profiles constitute a significant complement to the databank. The bill itself does not specify the assembly and employment process, the specification of

which the requirement of neutrality seems to foreclose. It is by examining this projected exercise that we shall approach our larger question. No comprehensive answer will be attempted. But the discretion appropriated to the DNA Profiling Board is abused as we shall see, in what they propose to do with the volunteered profile; and we shall risk saying thereupon, why that is not entirely surprising.

What Does the DNA Bill Do?

Let us begin with certain definitions the bill proffers in its first chapter. Here are the relevant ones:

(a) "DNA profile" means the result of analysis of a DNA sample that is intended to establish human identification, for instances listed in the schedule; (b) "DNA profiling" means a procedure to develop DNA profile for human identification, for instances listed in the schedule; (c) "DNA sample" means bodily substances of any nature collected for conducting DNA profiling, and the materials derived in a DNA laboratory, therefrom.

The circularity here looks harmless as (b) serves no purpose, actually, because (a) has linked "profiling" to "identification" already. The phrase "establish human identification" is to be glossed at as "establish human identity" presumably. Defining a "DNA sample" as DNA extracted from bodily substances, simply, would have obviated the need for (b) and avoided the circularity; doing so would have been more consistent with common usage, besides. The need to mention "materials derived" in (c) arises from the circumstance that the extracted DNA must be amplified, by the polymerase chain reaction typically in order to yield profiles. Profiling procedures may one day dispense with such amplification, so that neutrality does seem served by the generality of the statement here. But many sorts of "material," besides DNA, may be "derived" from bodily substances. The definition of "DNA sample" as it stands is much too broad, as any "result of analysis" performed on organic materials other than DNA, would now serve by the letter of the law as a DNA profile.

The drafting of the bill seems inept though, and not for some studied exploiting of the wish for technological neutrality, rather, by its proponents who do seem to be specialists—we should note now—in the relatively novel techne of DNA profiling. But we should be more circumspect perhaps. A human DNA profile could have been defined, quite simply and directly, as an identifier of a person derived from his or her DNA; after having defined an "identifier" first, if necessary, as anything that no two individuals are at all likely to share. The notion is not a legal novelty—fingerprints are routinely regarded so, after all—and one might wonder why the bill resorts to its circumlocutions. To proceed, we must consider the short tandem repeat (STR) profiles that have come into common use as profiles that will be stored in the DNA Data Bank, so we shall see. But we must recall some elementary facts first.

Composition of DNA Profiles

We need not consider, for our purposes, the precise physical and chemical structure of the DNA molecule, which may be summarised here as a sequence of just four nucleotides or bases—the names of which are abbreviated: A, C, G and T usually. The DNA replicated in each cell of a human individual may be termed as an individual genome. The eminently functional stretches within a genome are thought to be the coding sequences that determine the assembly of proteins in the cell—these sequences have come to be called genes. But it is very important that a gene is functionally identified first as a coding sequence for a specific protein, rather than a particular sequence of nucleotides. A typical "result of DNA analysis" would specify the bases constituting an identifiable coding sequence in an individual genome, as a string composed of the letters A, C, G and T—a "symbolic" artefact.

The genome of a human individual divides naturally into 23 chromosomal pairs. For any given pair, the father and the mother each contribute one chromosome, and the human genetic endowment is said to be diploid in consequence. A chromosome is a long stretch of DNA with a perduring physical and chemical structure. The constituents of a given chromosomal pair exhibit similarities of a structure that allows one to distinguish and isolate them, from the members of the other 22 chromosomal pairs. The chromosomes in a given pair are linked through their centromeres. The resulting alignment of complements allows one to identify the complementary regions on paired chromosomes—the loci of paired stretches of DNA mirror each other, or as one might say, be distinctive sequences of nucleotides that are much more like each other, in their particular succession, than runs of bases elsewhere on the chromosome. A distinctive run of bases, at an identifiable locus, may be a coding sequence for a specific protein—"functional stretches" of DNA do not extend over more than one chromosome, or so it is currently thought. Recall that a gene is functionally identified via a protein, and the coding sequences at a specific locus for a given protein may differ slightly across different individuals. The gene is said to have different alleles. Complementary loci are called "heterologous" when the paired alleles differ and, or are "homologous" otherwise.

With this very summary account we can return to the STR profile. The eponymous "short tandem" is a very short sequence of nucleotides, a string of some four or five bases typically. There are loci on chromosomes where the contiguous stretch of DNA consists of repeats of one and the same short tandem, which characterises the locus now. The number of repeats may vary across a complementary pair of loci. The count of repeats of a particular short tandem at a specified chromosomal locus has come to be called an STR-allele. The usage is awkward, for an allele is usually a sequence of bases and a "material given," rather than the symbolic artefact that a STR-allele has now become. But the departure from convention is innocuous, and only such counts could be "results of analysis" here anyway. For any actual stretch of repeated short tandems would be dispersed in the process of counting the repeats. To save writing, we need a term for the linked pair of such counts from any pair of complementary STR-loci, we shall call that an "STR-doublet" for

the nonce. So the "STR-allele" here is a positive integer and the "STR-doublet," a pair of such. Both are linked to a particular short tandem, characterising a specific STR-locus. A multi-locus STR profile now is a collection of STR-doublets, with each doublet linked to a complementary pair of loci, and though any set of such doublets is nominally a profile, it has become usual to reserve the term "multi-locus STR profile" for collections with enough paired STR-loci to securely identify human persons. Now the STR-loci, that are currently known, have between six and 18 different alleles each, and each allele, that is one element in an STR-doublet, would be a "result of analysis" already. But a single STR-allele is not a profile by any means and would serve to "establish" identity as coarsely as a blood group would; only there being so few at a given locus, the definition in the bill is defective for that reason alone.

DNA Identification Techne

We turn next to how many pairs of complementary loci an STR-profile would need in order to securely identify a person—a very important question. Let us term a complementary pair of STR-loci simply as "STR-position." Each STR-position will exhibit some finite number of STR-doublets in a given population "*P*," and each of these will have its own relative frequency, or proportion in that population. Now, the question of how many STR-positions might be needed in order to construct identifiers for individuals in *P*, has come to be settled in the following way: suppose L_1, L_2, ..., L_k are the STR-positions being used, with no two positions being on the same chromosomal pair. Let F_1, F_2, ..., F_k be the maximal proportions of the STR-doublets available at each location in *P*. Let "*Z*" be a particular STR profile assembled, using one STR-doublet from each of these k positions. And let Q_1, Q_2, ..., Q_k be the proportions of the k constituents of the profile in a given *P*. The product $\pi(Z) = Q_1, Q_2, ..., Q_k$ has come to be called the random match probability of *Z*, and the number is usually understood as the probability that a person if selected at random from *P*—and his or her profile assembled using just these STR-positions—then that profile would be identical to *Z*. The elementary procedure for computing $\pi(Z)$ has come to be called the Product Rule. We shall come back to this momentarily. Let #*P* be the size of the population *P*, and set $F = F_1, F_2, ..., F_k$. Note that $\pi(Z) \leq F$ then.

Suppose that the Product *F* is a minuscule quantity compared to the reciprocal 1/#*P* say, then $\pi(Z)$ will be minuscule as well. This circumstance is usually taken to ensure that were *X* the STR-profile of some person in *P*, then no other person in *P* is at all likely to precisely have the same *Z*—for a profile consisting of STR-doublets from the positions L_1, L_2, ... L_k used to assemble *Z*.

There is general confidence that STR-profiles will securely individuate persons, when their "random match probabilities" are thusly minuscule, and when these numbers have come to be presented in courts—especially in America and England—in order to secure probative force to matches of profiles. The proponents of the bill, we shall see, also share that confidence derived from taking the "Product Rule" to compute a probability. Now one may safely suppose that when the number of possible profiles is very small, compared to the size of the population, then the profiles will not be based on individuate persons of course.

But one cannot cogently suppose the rule to compute the probability of a random match, even if one assumes that profiles will be shared very rarely only (Mathews 2015). The arguments in Mathews (2015) are entirely elementary, but have been presented at great length and in numbing detail. We note, in passing, that a standard text on forensic DNA profiling (Thompson 2013) hesitates to regard the Product Rule as computing a probability, even though the author avers that the number, that the rule assigns a profile, will disclose its "rarity." Nonetheless, the reader who goes through Mathews (2015) will see that little is gained thereby. The judicial use of the Product Rule has been protested and a comprehensive questioning of its general application can be found in Taroni et al (2010). But whether the Product Rule computes a probability or not, or some "rarity" or the other, one can only tell whether the number it yields will be irrelevant (often enough) to deciding if an STR-profile does identify a particular person.

Suppose that the STR-positions L_1, L_2, ..., L_k are on some $j \leq k$ different chromosomal pairs. Let *Z* be above the profile of some actual person, who has a sibling of the same sex. The probability that the sibling has the same profile will equal or exceed $2/4^j$. This calculation is an elementary consequence of the fundamental thesis of an independent assortment in the formation of gametes. The exact probability depends on the number of homologous STR-positions on the parents' profiles—Section 4 of Mathews (2015) details how. The STR-loci are currently known to have between six and 18 different alleles, each of which we have noted. The number of distinct STR-doublets at a specific STR-position will vary between 62 and 182, therefore given the thesis of independent assortment, and the maximal proportions, F_1, F_2, ..., F_k will each be less than ¼th than we may safely guess, and very much less even.

So, for $j \leq k$ we will have:

(1) $2/4^j > 1/4^k >> F1, F2, ..., Fk \geq \pi(Z)$

Which shows, as we had asserted earlier, that the numbers computed with the Product Rule will be irrelevant to the question of whether, for a given person, a sibling of the same sex could have the same STR-profile as he or she does.

Let us pause to take stock. The bill defines a DNA profile as a "result of analysis" of a "DNA sample" and we have seen that the definition of a DNA sample is much too broad. Our consideration of multi-locus STR-profiles has disclosed this much: that the bill improperly construes the constituent integral alleles of such profiles as themselves. The DNA profile(s) for each such allele is a distinct "result of analysis" here, to note it again, and a single allele hardly serves to secure identity. The prime movers of the bill appear to be those that are in-charge of the centre for DNA Fingerprinting and Diagnostics, in Hyderabad—whose director had assured an experts committee that "the DNA profile of an individual may simply be described as a pair of numbers at each of the 17 neutral DNA positions," which taken altogether

"uniquely identifies" that individual. The envisaged profiles seem to be STR-profiles precisely. The "pair of numbers" here must be our STR-doublet, and the "DNA position" is what we termed as an STR-position.[1]

The supposition is confirmed when we find that the director of the Centre for DNA Fingerprinting and Diagnostics proceeded to assert that the corpus of volunteered profiles would be "absolutely essential to calculate the statistical probability values that would be needed in judicial forum, to establish that the match obtained between two profiles is not by chance alone." The expression, "calculate the statistical probability values," very likely means "estimate the probabilities" simply, and we are told immediately after that, "to calculate these values a one-time exercise is required, to obtain (with informed consent) the DNA profiles of about 100 random individuals from each of the different communities of India to be stored in the National DNA Data bank." That only "about a 100 profiles" are needed, makes it certain that the dubious Product Rule will be used to estimate the "statistical probability values" here. In the Appendix (p 68) we shall consider when and why these many STR-positions will suffice for that mistaken exercise, which may come as a surprise, initially, given that the quantum of STR-doublets at a good many positions may very well exceed a hundred. But it is clear now that our "expert" in "DNA Fingerprinting and Diagnostics" himself regards a DNA profile as an identifier of persons, which no STR-allele itself would be. So defining a DNA profile as the "result of DNA analysis" should have been unacceptably loose to the proponents of the bill.

Conclusions

The looseness of the definition here may or may not be a studied gambit, but the technological neutrality of the formula does abet, certainly, an appropriation of discretionary power to the specialists on the DNA Profiling Board. Whether the direct definition of a DNA profile that we have proposed would check that at all sufficiently is another matter of course. For the question, could a multi-locus STR profile serve as an identifier is a technical one; one that the law cannot easily address. There must be better ways of finding out, when 17 STR-positions are used, whether two persons could share the same STR-profile—two individuals who have not evolved from the same zygote, of course, for they will have the same genetic endowment. The director of the Centre for DNA Fingerprinting and Diagnostics does not seem to have looked into the matter. That he should resort to the Product Rule instead, is not surprising of course. But it is dismaying that the man should rely on it blindly without having examined the cogency of the calculation, simply because its use has been seconded by the US National Research Council—an "authoritative body of scientific experts" as our expert terms them—which has seen fit to declare that the Product Rule does "permit some reasonable estimate of how rare" an STR-profile is. The circumstance is sobering—given the elementary character of the adverse arguments we advance in Mathews (2015) especially. The expert practitioners of DNA profiling are less than diligent, it would seem, when they reason together upon probabilities. One must ask how prudent it is to concede any discretionary power, at all, to such specialists, particularly if the consequences of their decisions cannot be easily reversed.[2]

In the Appendix, we shall also state why the estimation of "statistical probability values" intended by our own experts in DNA Fingerprinting—using the envisaged corpus of volunteered profiles in the National Data Bank—would not always establish, in Indian "judicial fora," that "the match obtained between two profiles is not by chance alone." Were it not theoretically defective, even the Product Rule would fail to secure probative force in Indian courts—and would fail often enough. Why that is so, should be evident to the proponents of the bill themselves actually. The reason is not far to seek. The discretionary power that the bill concedes to the members of the DNA Profiling Board can be abused, we must conclude now, based on what they intend to do with the data bank.

Just how instructive is the case we have considered to the larger question, we began with, is itself a considerable question. That the technology in question is an "application" of a "life-science" is not a negligible thing. The circumstance that the mathematisation of biology is generally opportunistic, may matter very much now, and that the grant of discretion to specialists may present very little danger, all told, when a techne is grounded in a more fundamentally mathematised science like physics.[3] Legislation regulating the latter sorts of technology might be safely neutral then, though the matter may be harder to decide, altogether, when a techne has emerged from some decidedly computational and "simulative" way of conducting physical science.[4]

The regulation of "biotechnology" seems to present difficulties of another sort however: the imminent pervasion of daily life, by which technologies deriving from the science of genetics particularly, will soon demand comprehensive legislation. The Government of India proposes to institute a biotechnology regulatory authority—we should note now—and hard questions regarding the mathematisation of biology will be asked, one can only hope, before the inevitable concession of discretionary powers to the specialists, who will be authorised so.[5]

No technology of ours would have looked like τ ρκνη, one suspects to ancient Athenians. Were Plato to write his dialogues now, he would hesitate to have his master Socrates conjure these as exemplary foils to interlocutors, claiming such knowledge as he had particularly sought, the "project managers" of the engineering teams that design and assemble our smart-phones and killer-drones. To say why the philosopher would have hesitated, would need a treatise though, and that I should have introduced the word "techne," just as I did, may seem an egregious insinuation then. So it may be no excuse at all to record, now, that the question of whether medicine was a "techne" had already agitated the contemporaries of Socrates. And to note that the question turned on whether the "object" of the techne (health) was or was not a "definite" thing, that

admitted representation by "geometry" and "arithmetic."[6]

NOTES

1 I am quoting here from Item 2 in a "Record note of discussions of the Experts Committee meeting" held on 31 January 2013 at the Central Department of Biotechnology, in New Delhi, in order "to discuss potential privacy concerns on the draft Human DNA Profiling Bill." The epithet "neutral" was not elaborated upon.

2 The judicial consequences may be grave. The misleadingly minuscule numbers given by the Product Rule are cited to secure probative force to matches of STR-profiles in American and English courts. And how the elementary arguments of Mathews (2015) again induce us to agree with a recent text, on data analysis in forensic science, Taroni et al (2010), that the agonists in Anglophone judicial arenas are "stubbornly innumerate" indeed.

3 The difference between "opportunistic" and "fundamental" mathesis is indicated well enough in the prefaces to two texts: a comprehensive introduction titled *Mathematical Biology* (Murray 2002), which is in its third edition now, and a more recent book titled *Towards a Mathematical Theory of Complex Biological Systems* (Bianca and Bellomo 2011). The first text frankly subordinates the formal desiderata of mathematisation to the exigent demands of biology as it is practised, while the second seeks, in contrast, to orient biological research through a mathesis formally derived from thermodynamics.

4 "Application" and "theory" may be tangled inextricably together in such praxes, which people engaged in "science and technology studies" have come to call "techno-science."

5 The treatment of Gilles-Eric Seralini by the influential journal *Food and Chemical Toxicology* will have to be deeply considered, for that discloses how methodologically carious research on genetically modified organisms is. The lateral transfer of modifying transgenes from such organisms awaits some thorough mathesis as well, or some well-grounded simulation. The general risk is that such transfer poses cannot be properly assessed otherwise. The novel techniques of "Topological Data Analysis" by Carlsson (2009) may hold some promise here.

6 For a close discussion of Plato's "dialectical" resort to techne in Socratic dialogues one might look at David Roochnik's (1996) "Of Art and Wisdom."

REFERENCES

Butler, John (2009): *Fundamentals of Forensic DNA Typing*, Amsterdam: Academic Press.

Bianca, Carlo and Bellomo Nicola (2011*): Towards a Mathematical Theory of Complex Biological Systems*, Vol 11, Singapore: World Scientific.

Carlsson, Gunnar (2009): "Topology and Data," *Bulletin of the American Mathematical Society*, Vol 46, No 2, pp 255–308.

Mathews, Hans V (2015): "A Consideration of the Product Rule for STR-Profiles," Technical Report, Centre for Internet and Society.

Murray, J D (2002): *Mathematical Biology I: An Introduction*, third edition, Berlin, Germany: Springer-Verlag.

Roochnik, David (1996): "Of Art and Wisdom," *Plato's Understanding of Techne*, University Park Pennsylvania.

Taroni, Franco, Silvia Bozza, Alex Biedermann, Paolo Garbolino and Colin Aitken (2010): *Data Analysis in Forensic Science: A Bayesian Decision Perspective*, Chichester, West Sussex, UK: John Wiley & Sons.

Thompson, William (2013): "Forensic DNA Evidence: The Myth of Infallibility," *Genetic Explanations: Sense and Nonsense*, Sheldon Krimsky and Jeremy Gruber (eds), Cambridge, MA: Harvard University Press, pp 227–55.

Appendix

Let us consider the random sampling from "each of the different communities of India" that the proponents of the bill intend. Only "about a 100" or so individuals need to be picked out at random, they say, from each community. The purpose of the exercise must be to estimate the relative frequencies of STR-doublets in a given community, and in "judicial fora" the Product Rule would be used then presumably "to establish that the match obtained between two profiles is not by chance alone." By construing the simple product of these relative frequencies, for the doublets in the STR-profile of any given member of the community, we inch closer to the chance of its matching the profile of any person randomly selected from the same community. Doing so is not licit. And the reader, who is keen to know why, can refer to Mathews (2015) again.

But setting aside the soundness of the Product Rule, let us consider why only a 100 or so individuals need to be sampled from a given community, for that would seem too few, given the number of doublets an STR-position could have. But just that many would do, if the relative frequencies of STR-doublets could be calculated from the relative frequencies of the few STR-alleles that give rise to them, of which there are between six and 18 only, at each known STR-position. Now doublet frequencies can indeed be calculated from allele frequencies, at a given position, if the alleles there are in the Hardy–Weinberg equilibrium. It is an elementary result in population genetics that the alleles will very soon come to be so in a randomly mating endogamous population. So each "community" is now supposed to be an endogamous population, in which males and females couple "at random" as husbands and wives.

The absence of social constraints on marriage is not the supposition. What is being assumed is that the short tandems repeats, being selectively neutral, would play no role in determining who marries whom. But the complication now is this, sound or not, the Product Rule cannot be used on the profiles of persons whose parents belong to different endogamous communities. The use of STR-profiles in criminal trials would be stymied for such defendants, therefore, if the Product Rule applied to doublet frequencies it would give them their judicial utility and such "miscegenation's" happen often enough surely.