at presentation stage and, second, no statistician has yet to suggest to me a more practical method of solving the problem. I eagerly await the response of readers of this journal.

Two of the other discussants raise the issue of band independence and multiplying probabilities. This has attracted enormous interest in the United States. A lot of it, apparently, from eminent people who hitherto had little or no experience in the forensic field. It is good that such people should take an interest, but it is important that they should realize that they have entered a field that may have requirements, strictures and problems that are different from those that they are accustomed to working with. Frankly, I find the discussions of Hardy–Weinberg equilibrium, linkage disequilibrium and population substructuring confusing and often of no more than tangential interest. The important question for me is not "can I prove independence?" but "is there any evidence of dependence effects that would have any practical impact on operational casework?" The robustness studies that I have carried out with colleagues have served to increase the confidence with which we carry out our casework procedures. Even when we have constructed artificially stratified simulated populations (Evett and Gill, 1991), we have failed to produce effects which would cause operational disquiet. We have carried out all $N(N-1)/2$ comparisons in a file of $N$ Caucasians using a file of Afro-Caribbeans for estimating frequencies (Evett and Pinchin, 1991) and have shown that even under these conditions our operational procedures are robust. Our most recent work (Evett, Scranage, Pinchin and Buffery, 1991) has shown that, if there are any between-probe dependence effects in U.K. Caucasians, they are too small to have any practical effects in case work. We do not take these results as a source of complacency, nor do we claim that they have universal implications for all countries and racial groups. However, we do suggest that the fears that have been expressed may sometimes grow out of reasonable proportion.

# Rejoinder

**Donald A. Berry**

I thank the dicussants for their clear and insightful comments. All discussants have important concerns and identify important problems for future research. I am pleased that all four seem to favor the approach I describe in preference to match/binning. The editors tried to find discussants who use match/binning and who would argue its merits, but unfortunately they were unsuccessful.

While I have no major disagreements with the discussants, I will respond to some of the points they raise.

## RESPONSE TO LANGE

Lange correctly points out that I did not dwell on departures from the independence assumptions. In Berry, Evett and Pinchin (1991), we extend the results of the current paper to the bivariate setting of pairs of bands on a single-locus probe. The approach does not assume independence of the two bands. The second "key independence assumption" is more difficult to relax since going to higher dimensions has calculational and sample size implications. Unpublished results of Evett and his colleagues (see Evett's discussion) indicate that independence is not a concern for the probes used in the UK Home Office Forensic Science Service. However, *measurement errors* across probes are highly correlated; research to account for such "band shifting" across probes is ongoing.

Lange likes the name "identity index" for R. I like it too. Actually, while I will continue to use both, neither "Bayes factor" nor "likelihood ratio" is ideal. The former carries a bit more philosophical baggage than R deserves. The latter is somewhat of a misnomer because R involves Bayesian averaging in both numerator and denominator.

Lange worries about the ability of judges and juries to adjust priors to posteriors. This worry is shared by Kaye, Evett and many others—including me! Judges and juries should be given (1) information they can understand, and (2) information that is correct. There can be no compromise regarding (2). If something we provide is correct but we know it is likely to be misunderstood, then we may be clean legally but we have failed intellectually.

Presenting a Bayes factor (or a match/binning proportion) to a court without instructions on interpreting it is a case in point. I recently had an exchange with a writer for a scientific publication who wanted to report the Bayes factor as $P(G \mid E)/P(I \mid E)$ rather than $P(E \mid G)/P(E \mid I)$. Apparently, she wanted to have a posterior without having a prior—hardly the first attempt at that! When I protested, she replied that her readers were scientists and not statisticians and therefore few would understand or appreciate the subtle difference. I agreed but convinced her not to say something she knew to be wrong. (Actually, she had no idea whether it was wrong—she knew only that I said it was wrong.)

Confusing likelihoods with posterior probabilities is common and, I think, inevitable. Jurors cannot help but view a Bayes' factor (or a match/binning proportion) as posterior odds (cf. Evett's discussion). At a minimum they should be told that this view is incorrect and that it ignores the other evidence in the case. As I indicated in my main article, I think we should go further and try to show the jury how to incorporate the other evidence and come up with a juror-specific probability of guilt. But despite the tone of my article, I am not convinced that this would do more good than harm.

I agree with Lange that PCRs have benefits over RFLPs. I also agree that the statistics involved in using PCRs is less interesting—and that a simpler inferential process is one of the benefits of PCR technology. But there are many interesting statistical issues in PCR technology. In particular, the measurement error problem may be alleviated somewhat, but I think it will not go away. And I do not agree with Lange that checking for possible linkages among the many loci used is easy.

## RESPONSE TO CHERNOFF

Chernoff claims that a basic flaw in the match/binning approach is multiplying probabilities. Obviously, I am no fan of match/binning. But —except for the possible lack of independence—I do not see multiplying probabilities as a flaw. Chernoff points out that I did not make it clear whether match/binning probabilities are $P$ values. I apologize for this oversight. They are not $P$ values; match/binning does not involve calculating tail areas. So some of Chernoff's remarks that deal with $P$ values do not apply to match/binning.

I am concerned that I did not adequately describe match/binning. Here's a way of thinking about it. Suppose humans could be characterized on the basis of several discrete systems (such as genetic systems like ABO blood type, which have a finite number of alleles). And suppose these characteristics are known for a certain criminal. A suspect is found whose characteristics perfectly "match" those of the criminal. It makes sense to ask what proportion of the population similarly matches. And in calculating this proportion it is correct to multiply proportions of the individual systems—if they are independent. So far, so good. Match/binners first discretize an essentially continuous system and then proceed as though it actually is discrete. My objections do not concern the way they do this, but the fact that they do it at all.

I agree with Chernoff that my smoothing techniques are ad hoc. I am comfortable taking $b = 1$ to account for measurement error. However, increasing $b$ beyond 1 to account for sampling variability is but one of many possible approaches, some of which I have indicated, and all of which come with some level of ad hockery.

As Chernoff indicates, my smoothing techniques lower the largest observed frequencies—in essence robbing the rich to give to the poor. I think this is good on the whole, but it obviously has some bad aspects. It increases $R$ when the observed band weights are very common, and so it is not conservative. Evett has suggested to me the possibility of smoothing without decreasing any observed frequencies. This would be conservative, but it also means having total density greater than 1. This notion repelled me at first, but I now think this was an overreaction and was due to one of my axiomatic hang-ups!

In his section on *Reference Population*, Chernoff suggests that it usually makes sense to use the population to which the suspect belongs. I agree that this is conservative *on the average*. But we don't have to average: we know the measured band weights. Suppose a suspect has a band weight that is very rare for his race but common for other races. Using his race for reference gives a much larger Bayes factor $R$ (and a much smaller match proportion) than would using other races. Should we use the race corresponding to the smallest R? Concerning Chernoff's statement, "In the Ponce case the issue was whether the blood belonged to the victim, and there the natural reference population is that of the victim," I don't see what's natural about it. Consider a match proportion. This is the proportion of the observed bands assuming the blood on Castro's watch is *not* that of the victim. So the question is: If it's not the victim's, whose is it? That person may or may not be Hispanic.

In the section on *Guilt versus Matching*, Chernoff says that "Matching of the DNA strands may not imply guilt." The converse is also true. As I indicated on page 186, "In this example the suspect [Castro] may be innocent even if the blood on his watch is that of the victim. On the other hand, he may be guilty even if the blood on his watch is not that of the victim. Handling these possibilities is easy, but I will avoid the extra algebra by assuming P(blood on watch is Ponce's | $G$) = 1 and $P$(blood on watch is Ponce's | $I$) = 0."

I do not understand Chernoff's point (in 6.B) that "It should be remarked that if Lifecodes type bins are used in a nonexclusionary fashion, apparent matching on almost all of the highly polymorphic probes could be strong evidence, even if there were apparent failure to match on one or two probes." If the "apparent failure" can be explained—by the possibility, for example, that there are missing bands—then I agree. But if there is a clear disagreement between any single band measurement in the suspect and crime samples, then they cannot have come from the same individual.

Chernoff's last sentence makes a very good point, one I have not seen before. Suppose—as is not unusual—the crime sample is tiny and so will allow testing with only a small number of probes. Which probes to use? Since the suspect's sample is usually ample, test it first using all available probes. Condition on the suspect's observed band weights and choose the probes that will be most informative; usually these will be loci for which the suspect has rare alleles. Using the probes chosen, compare suspect and crime samples by analyzing them together on the same autoradiogram.

## RESPONSE TO KAYE

Kaye gives an elegant account of problems using probability, in the Bayesian approach in particular, in court. I find nothing to argue with in what he says, but I will comment on some of the issues he raises.

I like Kaye's "weak" versus "strong" Bayesian dichotomy, but I suggest that the names could be improved: Would anyone admit to being a "weak Bayesian"? "Limited" and "full" are hardly better. Perhaps "first degree" and "second degree"? Also, it might be appropriate to have a name for the third category Kaye describes, the one that is weakest of all: present only the Bayes factor R. As I indicated in response to Lange, I would like to go beyond this weakest level, but I would be satisfied to stop at "weak."

I agree that the contamination and degradation of samples are important problems. Suspect samples should be of good quality and so these problems should occur mainly with crime samples. Degradation can result in missing bands, and the larger the band, the more likely it will be missing. Also, very small bands can "run off the end" because of process and measurement limitations. Berry, Evett and Pinchin (1991) address missing bands from a Bayesian perspective using subjective assessments of the probabilities involved. Contamination that may introduce additional bands can be handled similarly.

## RESPONSE TO EVETT

I wrote the article being discussed almost two years ago. Since that time I have collaborated with Evett on inferential aspects of DNA profiling, and he has taught me much. That continues to be the case in his discussion. He highlights some of the changes in attitudes toward statistical inference that have taken place among forensic scientists in recent years, and it is refreshing to see that Bayesian ideas are becoming increasing important. In addition, Evett's discussion of the difficulties in presenting quantitative evidence to courts is intelligent and thought provoking.

I do not favor Evett's "verbal convention," although I do not consider it "rank heresy." At best, it makes incorporating prior opinions difficult, and I believe such opinions are essential for interpreting evidence. At worst, a verbal convention might be taken by jurors as a posterior probability assessment, even though this is clearly not intended. For example, Evett says that $R = 450$ is "strong evidence." If the suspect had been selected in some arbitrary fashion from a large population, my posterior probability of guilt could easily be quite small. But if the suspect had been arrested at the scene of the crime, then my posterior probability of guilt would be rather large.

## ADDITIONAL REFERENCES

BUDOWLE, B., CHAKRABORTY, R., GIUSTI, A. M., EISENBERG, A. J. and ALLEN, R. C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *American Journal of Human Genetics* **48** 137–144.

*Caldwell v. State*, 260 Ga. 278, 393 S.E. 2d 436 (1990).

CHERNOFF, H. (1979). *Sequential Analysis and Optimal Design*, rev. ed. CMBS Series in Applied Mathematics, SIAM, Philadelphia.

COHEN, L. J. (1981). Subjective probability and the paradox of the gatecrasher. *Arizona State Law Journal* **1981** 627–634.

DEVLIN, B., RISCH, N. and ROEDER, K. (1990). No excess of homozygosity at loci used for DNA fingerprinting. *Science* **249** 1416–1420.

DEVLIN, B., RISCH, N. and ROEDER, K. (1991). Forensic inference from DNA fingerprinting. Unpublished manuscript.

EVETT, I.W. (1987). On meaningful questions: A two trace transfer problem. *Journal of the Forensic Science Society* **27** 375-381.

Evett, I. W. and Gill, P. (1991). A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations. *Electrophoresis* **12** 226-230.

Evett, I. W. and Pinchin, R. (1991). DNA single locus profiles: Tests for robustness of statistical procedures within the context of forensic science. *Human Genetics*. To appear.

Evett, I. W., Scranage, J. K. Pinchin, R. and Buffery, C. (1991). A test for independence between DNA RFLP probes in forensic casework. Unpublished manuscript.

Fienberg, S. E. and Kadane, J. B. (1983). The presentation of Bayesian statistical analyses in legal proceedings. *The Statistician* **32** 88-98.

Geisser, S. (1990). Some remarks on DNA fingerprinting. *Chance: New Directions for Statistics and Computing* **3** 8-9.

Kadane, J. B. (1990). A statistical analysis of adverse impact of employer decisions. *J. Amer. Statist. Assoc.* **85** 925-933.

Kaye, D. H. (1987). The polygraph and the PVP, comment on "The statistical precision of medical screening procedures: Application to polygraph and AIDS antibodies test data" by J. L. Gastwirth. *Statist. Sci.* **2** 223-226.

Kaye, D. H. (1988a). What is Bayesianism? In *Probability and Inference in the Law of Evidence* (P. Tillers and E. Green, eds.) 1-19. Kluwer, Hingham, Mass. (Reprinted in 1988, *Jurimetrics* **28** 161-177.)

Kaye, D. H. (1988b). Plemel as a primer on proving paternity. *Willamette Law Journal* **23** 867-883.

Kaye, D. H. (1989). The probability of an ultimate issue: The strange cases of paternity testing. *Iowa Law Review* **75** 75-109.

Kaye, D. H. (1990a). DNA paternity probabilities. *Family Law Quarterly* **24** 279-304.

Kaye, D. H. (1990b). Presumptions, probability and paternity. *Jurimetrics Journal* **30** 323-349.

Kaye, D. H. (1991). Neither sheep nor trumpets: Statistics for lawyers and law for statisticians. *Michigan Law Review*. To appear.

Kaye, D. H. and Kanwischer, R. (1988). Admissibility of genetic testing in paternity litigation: A survey of state statutes. *Family Law Quarterly* **22** 109-115.

Lenth, R. V. (1986). On identification by probabilty. *Journal of the Forensic Science Society* **2641 197-213.**

Li, H., Gyllensten, U., Cui, X., Saiki, R., Erlich, H. and Arnheim, N. (1988). Amplification and analysis of DNA sequences in single human sperm. *Nature* **335** 414-417.

Litt, M. and Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within a cardiac muscle actin gene. *American Journal of Human Genetics* **44** 397-401.

McNally, L., Baird, M., McElfresh, K., Eisenberg, A. and Balazs, I. (1990). Increased migration rate observed in DNA from evidentiary material precludes the use of sample mixing to resolve forensic cases of identity. *Applied and Theoretical Electrophoresis*. **1** 267-272.

Meier, P. and Zabell, S. (1980). Benjamin Pierce and the Howland Will. *J. Amer. Statist. Assoc.* **75** 497-506.

Nesson, C. (1985). The evidence or the event? On judicial proof and the acceptability of verdicts. *Harvard Law Review* **98** 1357-1392.

Office of Technology Assessment (1990). Genetic witness: Forensic uses of DNA tests. Government Printing Office, Washington, D.C.

*People v. Castro*, 144 Misc. 2d 956, 545 N.Y.S. 2d 965 (Bronx County Sup. Ct. 1989).

Robbins, H. E. (1968). Estimating the total probability of the unobserved outcome of an experiment. *Ann. Math. Statist.* **39** 256-257.

*Robinson v. Mandell*, 20 F. 1027 (C.C.D. Mass. 1868).

Shaviro, D. (1989). Statistical-probability evidence and the appearance of justice. *Harvard Law Review* **103** 53-554.

*State v. Carlson*, 267 N.W. 2d 170 (Minn. 1978).

*State v. Kim*, 398 N.W. 2d 170 (Minn. 1987).

*State v. Klindt*, 389 N.W. 2d 670 (1986).

*State v. Schwartz*, 447 N.W. 2d 422 (Minn. 1989).

Thompson, W. C. and Ford, S. (1991). The meaning of a match: Sources of ambiguity in the interpretation of DNA prints. In *Forensic DNA Technology* (M. A. Farley and J. J. Harrington, eds.) 93-152. Lewis, Chelsea, Mich.

Tribe, L. (1981). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review* **84** 1329-1393.

*United States v. Two Bulls*, 918 F.2d 56 (8th Cir. 1990).

*United States v. Jakobetz*, 747 F. Suppl. 250 (D. Vt. 1990).

Weber, J. L. and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44** 388-396.