JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, November 2009, Vol. 54, No. 6
doi: 10.1111/j.1556-4029.2009.01191.x
Available online at: interscience.wiley.com

Commentary on: Authors' Response [J Forensic Sci 2009; 54(2):501] to Wells' comments [J Forensic Sci 2009;54(2):500] regarding Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, Kornfield IL, Risinger DM, Rudin N, Taylor MS, Thompson WC. Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. J Forensic Sci 2008;53(4):1006–7.

Sir,

Let me say at the outset that I agree with the call by both Wells and Krane et al. for additional research into potential observer effects and/or bias in decision-making by forensic examiners. However, in my opinion the situation is not nearly as clear as Krane et al. suggest in their response to Dr. Wells. In that response, eight studies are cited in the statement "Empirical studies have confirmed that observer effects can influence [the results in various forensic disciplines]." My first impression upon reading this was that these studies showed that observer effects and/or bias were "proven" to be concerns in the listed disciplines.

But one reference, in particular, caught my eye. The 1984 Miller (1) study may be familiar to some as it has been mentioned at least twice before; once at the 2008 AAFS meeting in Washington, DC (2) and again at a recent "Expert Forensic Evidence" conference in Toronto, ON (3). Despite its provocative title, this study provides absolutely no data that could be construed even remotely as pertaining to qualified forensic document examiners. Rather, "Twelve college students, trained in the forensic examination of questioned documents, were utilized in the experiment" (p. 409). No details were provided by the author about the nature of the training given to the students but I find it difficult to understand how the results of a study based entirely on college students can be extended to professional, qualified examiners in any meaningful manner.

Upon seeing this reference in the list of citations I decided to review all of the studies. My review showed that these studies provide limited, even ambiguous, data with respect to whether or not observer effects or bias are issues with qualified forensic examiners. Indeed, like the Miller study described above, three of the cited studies did not involve professional forensic examiners at all and instead used only students as test subjects. One study was a meta-analysis based upon two earlier studies. The remaining three studies provide rather conflicting data about the potential for bias/observer effects. I would encourage everyone to review these articles for themselves but a short discussion of each is provided here to clarify my position on this matter.

Both of Miller's studies from 1984 and 1987 (4) used students and included no trained examiners. In the 1987 study the author stated "Fourteen students enrolled in advanced crime laboratory college courses were selectively trained in human hair identification techniques. The training consisted of 60 academic hours of lecture and 60 academic hours of laboratory experience under the instruction of court-qualified human hair experts. The 14 students met the basic requirements for expert testimony on human hair identification in courts of law. Each of the 14 examiners was independently advised to examine and compare human hair evidence in four criminal investigations" (p. 160). The author did not say if the students had successfully completed their courses; only that they were enrolled in such courses. The author's assertion that the students "met the basic requirements for expert testimony" is open to interpretation. As most readers know, meeting the basic requirements to be qualified as a court expert may not, in fact, make someone qualified to do the work. Beyond this, even if the college students were considered to be advanced trainees (e.g., novice examiners) rather than naïve subjects, the extension of these results to fully qualified forensic examiners is dubious.

The Beckham et al. study (5) in 1989 was quite extensive in that it involved 180 mental health experts who were asked to assess NGRI (Not Guilty by Reason of Insanity) submissions. Beckham et al. commented in part that "in the current study, no statistically significant bias was detected between groups…" (p. 86). In their discussion of this finding the authors further commented "Even though such bias has been demonstrated in clinical psychology graduate students, practicing forensic evaluators may be more attuned to such detrimental possibilities and therefore actively strive to be as objective as they can" (p. 86). If anything, this particular study suggests that bias was less of a problem than the researchers had anticipated though, of course, there is the issue of extension of the results to other types of forensic work. At any rate, it certainly does not support the belief that observer/bias effects are present in all types of forensic work.

The Dror et al. study (6) in 2005 involved "…27 university student volunteers, with a mean age of 23 (9 were males and 18 were females)." In their discussion of the results, the authors noted this limitation and commented "Second, our findings need to be examined within the context of routine everyday work of fingerprint experts. The training, experience, and work procedures of fingerprint experts may play an interesting and crucial role in if and how top-down components play a role in fingerprint identification. On the one hand, fingerprint experts may be less susceptible to top-down interference, perhaps even immune, to such effects. Given their highly specialized skills, they may be able to focus solely on the bottom-up component and be data driven without the external influences that we have observed in the research reported here. On the other hand, and in contrast, fingerprint experts may be even more susceptible to such top-down components" (pp. 807–808). Overall, these comments suggest to me a rather inconclusive position; a very reasonable position since the data from the study had nothing to do with qualified examiners.

Following the 2005 study, two studies (7,8) from 2006 by Dror et al. used qualified fingerprint experts (five and six, respectively). These studies are arguably the most intriguing to date insofar as they provide some support for the belief that observer effects may influence fingerprint examiners in at least some situations. At the same time, the nature of the influence/bias effect is not entirely clear. Aside from the issue of generalization of results from relatively small sample sizes (a point discussed by the original authors), the bias effect seems to be mostly unidirectional. That is, bias attempts may shift some conclusions toward exclusion but they were not very successful in moving conclusions toward individualization. In their *Journal of Forensic Identification* article, the authors speculated about this result saying "It seems that the threshold to make a decision of exclusion is lower than that to make a decision of individualization. Indeed our data support this claim, as reflected by the fact that most of the conflicting decisions were past individualizations. We did, however, observe a case in which an exclusion decision was now judged to be an individualization. This relates to the decision-making model used by experts in the fingerprint domain" (p. 613). In the end, I would agree there is evidence to support the belief that observer effects can be a factor in these types of comparisons but the precise nature, and the limits, of the influence is not clear from these studies.

Schiffer and Champod's study (9) in 2007 did not use fully trained examiners. Their test subjects were 48 students in forensic sciences studies. How the education of these students might compare with that in the 1987 Miller study is unknown, but the domains were clearly different. With respect to possible bias effects, the authors wrote "Contrary to our initial expectations for test II on the potential effects of stimuli inducing observational biases, no effect of availability of known print nor context information has been observed. This was true for all fingermarks used in the test. These results do, to a certain degree, contradict previous findings or hypotheses, for instance Risinger et al. (10) and their overview of studies on the detrimental effects of expectation on reasoning and perception" (p. 119). The authors pointed out limitations in their study. It is interesting that the findings did not support the idea that observer effects/bias are a problem with this type of examination. However, the key issue with this study is the same as with the other student-based efforts; namely, how well would these results extend to fully qualified forensic examiners?

Dror and Rosenthal's 2008 study (11) was a meta-analysis of the two 2006 studies (discussed above) and, as such, does not provide additional empirical support beyond those studies. This study involved an analysis intended to clarify the strength of the results from the earlier studies which were limited to a small number of subjects. As such, the authors concluded "The first two studies to examine these questions established that experts are far from being perfect. These studies demonstrated circumstances in which experts were both relatively unreliable and biasable, and in the analyses reported here we quantify these effects statistically and subject them to meta-analytic procedures. The data are based on forensic decision-making made by latent fingerprint experts, but because this forensic domain is the most widely used and well established, we can be confident that the problems exposed within this domain are also prevalent in other forensic domains" (p. 903). In my opinion, there is no reason to consider this particular type of work to be representative of other forensic domains simply because it is widespread and well established. Nonetheless, the idea that bias effects are present in at least some situations has support according to the authors. At the same time, they also commented, "The fact that fingerprint experts can be unreliable and biasable does not mean that they are not ordinarily reliable and unbiasable" (p. 903).

In general, the use of students as test subjects, whether they be completely naïve students or novice examiners, is inappropriate if the intent is to learn about the behavior of fully qualified examiners. Once studies based on students are removed from consideration, the remaining works are not conclusive one way or the other. Indeed, the two studies by Dror et al. in 2006 and the 1989 Beckham et al. study provide conflicting information. The results suggest, on the one hand, that contextual information can selectively bias the results for fingerprint examiners while, on the other hand, mental health experts did not seem to be subject to a significant bias effect at all. There may be a number of reasons why the results of these studies diverged so much including, for example, the different domains under consideration or the very different experimental designs. But the bottom line is that evidence regarding the existence or impact of observer effects or bias cannot be considered conclusive or even consistent.

Therefore, I have to agree with Dr. Wells in his assessment of the situation when he wrote "the incidence of such bias is unknown." In their response to Dr. Wells, Krane et al. suggested that the issue of observer effects has been confirmed in several other forensic domains and, based on this, it is only logical that it will be present in DNA analyses as well. Yet in reality there is no clear empirical evidence to support the belief that this is a problem in the other disciplines, let alone for DNA interpretations. Perhaps there is a problem and perhaps there isn't.

I personally believe there are grounds to warrant well-designed research studies aimed at gaining a better understanding of the situation. Indeed, I would support this based solely on the belief that "observer effects are a basic phenomenon of human psychology." As such, these issues may ultimately prove to be important factors in forensic decision-making. Or they may turn out to be relatively meaningless.

The implementation of any solution before the "problem" is fully understood is not a good idea. That approach may well result in new and unanticipated issues that end up being worse than the original concern. Let us do the research and understand the situation more fully before we begin "fixing" things that may or may not need to be fixed.

References

1. Miller LS. Bias among forensic document examiners: a need for procedural change. J Police Sci Admin 1984;12:407.
2. Risinger DM. The impact of confirmational bias and context effect on report writing in the forensic science laboratory. Proceedings of the 60th Annual Meeting of the American Academy of Forensic Sciences, February 18–23, 2008; Washington, DC. Colorado Springs, CO: American Academy of Forensic Sciences, 2008.
3. Saks MJ. Building an evidence-based report accounting for confirmation bias. Presented at the Expert Forensic Evidence in Criminal Proceedings: Avoiding Wrongful Convictions Conference, May 9, 2009; Toronto, ON. Toronto, ON: Osgoode Hall Law School of York University and the Centre for Forensic Science & Medicine at the University of Toronto, 2009.
4. Miller LS. Procedural bias in forensic examination of hair. Law Hum Behav 1987;11(2):157–63.
5. Beckham JC, Annis LV, Gustafson DJ. Decision making and examiner bias in forensic expert recommendations for not guilty by reason of insanity. Law Hum Behav 1989;13(1):79–87.
6. Dror IE, Peron A, Hind SL, Charlton D. When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints. Appl Cogn Psychol 2005;19(6):799–809.
7. Dror IE, Charlton D, Peron A. Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 2006;156:74–8.
8. Dror IE, Charlton D. Why experts make errors. J Forensic Ident 2006;56(4):600–16.
9. Schiffer B, Champod C. The potential (negative) influence of observational biases at the analysis stage of fingermark individualization. Forensic Sci Int 2007;167:116–20.
10. Referring to: Risinger MD, Saks MJ, Thompson WC, Rosenthal R. The Daubert/Kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion. Calif Law Rev 2002;90:1–56.
11. Dror IE, Rosenthal R. Meta-analytically quantifying the reliability and biasability of forensic experts. J Forensic Sci 2008;53(4):900–3.

Brent Ostrum,[1] B.A. (Hons)
[1]Senior Forensic Document Examiner, Canada Border Services Agency, 79 Bentley Avenue, Ottawa, ON, Canada K2E 6T7.