

## Authors' Response

Sir,

We appreciate the careful attention Brent Ostrum (1) has given to our recent letters and his close examination of the studies we have cited in support of the hypothesis that observer effects are both real and important in forensic science. Ostrum calls for more research on observer effects, as do Wells (2) and the recent NAS report (3). We agree that more research is desirable and view these calls as a favorable development in forensic science. However, we disagree with the suggestion that it is somehow prudent to postpone implementing sequential unmasking type safeguards until we have accumulated more data on observer effects in forensic science. Before explaining why we consider such a posture imprudent, we explain why we think the existing evidence is stronger than Ostrum apparently allows.

Ostrum states that Miller's 1984 study (4) "provides absolutely no data that could be construed even remotely as pertaining to qualified forensic document examiners." He bases this judgment on the fact that Miller describes his experimental subjects as "college students." Risinger (5), however, reports information obtained from Miller (6) that is relevant to this criticism. While Miller's 12 subjects were all part-time college students, four of them were "court-qualified document examiners working for police agencies" and the remaining eight "had completed training but had not yet testified in court." These 12 examiners were divided into two groups of six. Group I was given the unknown samples (three forged checks) and exemplars from one known source and were exposed to potentially biasing information before examining the evidence. Group II was given the same unknown samples and known samples from three sources and were shielded from any potentially biasing case information. "Four examiners in Group I, including one of the court-qualified examiners, incorrectly concluded that the suspect (as represented by the known exemplar) wrote the signatures on the three checks. The other court-qualified examiner declared the results of his examination to be 'inconclusive,' asserting that the known exemplars of the suspect's handwriting 'bore disguised handwriting characteristics.' The last examiner (a trainee) correctly eliminated the suspect. All six examiners in Group II correctly eliminated all three of their suspects." An inexperienced, though fully trained, examiner in Group I outperformed both court-qualified examiners.

Ostrum also states that Miller's 1987 study (6) "used students and included no trained examiners." He views Miller's statement that they "met the basic requirements for expert testimony" as "open to interpretation" and remarks that "meeting the basic requirements to be qualified as a court expert may not, in fact, make someone qualified to do the work." Ostrum views the "extension" of Miller's 1987 results to "fully qualified forensic examiners" as "dubious." Risinger has obtained from Miller (personal communication, 2009) further information regarding the subjects of his 1987 study and finds that Miller's subjects were again more qualified than Ostrum believes. The test subjects were, as before, students at East Tennessee State, but all had completed training in visual hair analysis under instructors who were qualified to give such instruction professionally, and whose trainees either currently worked in law enforcement or were regularly hired by law enforcement laboratories. While Professor Miller is unsure whether any of those subjects had testified at the time of the study,

in the normal course of events they would have been testifying to their results in court within a short period of time.

Ostrum goes on to say that the 2006 studies by Dror et al. (7,8) "provide some support for the belief that observer effects may influence fingerprint examiners in at least some situations." He says, "the nature of the influence/bias effect," however, "is not entirely clear." Passing over the question of sample size, Ostrum comes to his main point in this regard: "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." It is true that the Dror and Charlton study (7) demonstrated more switching from inclusion to exclusion than *vice versa*. Their subjects had 24 opportunities to switch away from individualization and 24 opportunities to switch away from exclusion. Their subjects switched from individualization five of 24 times, four to exclusion and once to "cannot decide." Their subjects switched away from exclusion, however, only one of 24 times, that one being a switch to inclusion.

This difference in effect size may suggest that fingerprint examiners tend to set their decision thresholds higher for individualization than for exclusion. Even if they do, four considerations suggest that we should give more weight to the one switch to inclusion than Ostrum seems to have done.

First, in many forensic contexts, biasing stimuli are often in the direction of individualization. Evidence is more likely to be submitted when it is thought to be incriminating rather than exculpatory, thus tending to create a bias toward individualization even when other domain-irrelevant information is absent. In a detailed study of four different crime laboratories (8), on average, greater than 90% connected a suspect with a crime scene or to the victim. This high rate of inculcation may come from the fact that each piece of evidence connected with a suspect *a priori* has a greater likelihood of being inculpatory. Thus, the false individualization rate in practice may be higher than the results of Dror and Charlton (7) seem to suggest.

Second, false exclusions are also undesirable errors in that they may let the guilty go free. Even if observer effects were somehow shown to produce only false exclusions, it would be appropriate to adopt sequential unmasking to minimize such errors.

Third, Wertheim et al. (9) recently found that, at least in some situations, bias may lead mostly to inconclusives. While this may seem neutral and harmless on its face, "inconclusives" have the potential to mutate to "can't exclude" in court testimony directed by a persistent prosecutor.

Finally, even if we accept that the size of the effect observed by Dror and Charlton (7) is representative of the size of the effect in practice, there is at a false-positive rate of at least 1 in 48 (roughly 2%) which is surely high enough to warrant a preventive response.

We also give more weight to the other studies we have cited than Ostrum does regarding the question of whether training and experience reduce one's susceptibility to observer effects. Ostrum seems to feel that training and experience are curative when he says, "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." If training and experience were curative, it raises the alarming prospect that each examiner has an initial period of practice in which he or she may be providing unreliable analysis and testimony. Such

a possibility would seem to be an argument in *support* of sequential unmasking.

Because susceptibility to observer effects is a human universal, it is implausible to suggest that experience would somehow eliminate all susceptibility to it. Little evidence exists that de-biasing strategies work well in forensic science practice. Even if the effect size were small, it is better to minimize it.

The notion that experience as a forensic scientist could cause one to transcend observer effects is akin to the notion that experience as a pilot or painter could cause one to transcend color blindness. In both cases the infirmity is built into the human architecture and cannot be willed away. The main difference is that observer effects are universal, while color blindness afflicts only a fraction of the population. We saw earlier that a trainee outperformed experienced examiners in identification (4). Statistical analysis of the various CTS handwriting proficiency tests by the CTS itself has never shown accuracy to be a function of years in practice. The Miller study could even be read to suggest that field experience may make one *more* susceptible to observer effects. Several mechanisms suggest themselves. First, experience produces routine, and routine reduces one's alertness to possible errors (10). In other words, one grows complacent, and perhaps more susceptible to subtle context cues. Second, interacting primarily with law enforcement can limit the type of information to which a forensic scientist is exposed. As the recent NAS (3) study notes, "Forensic scientists who sit administratively in law enforcement agencies or prosecutors' offices, or who are hired by those units, are subject to a general risk of bias." Finally, experience may reduce or eliminate the nervous novice tendency to self-doubt, thus tending to increase overconfidence among experienced examiners. Just as most drivers report their skills are better than average (11), most experienced forensic scientists may consider themselves less likely than average to commit an error. While these considerations are speculative, they shift the burden of proof onto those claiming that experience reduces susceptibility to observer effects.

We said earlier that we do not accept Ostrum's suggestion that it is somehow prudent to postpone sequential unmasking until we have accumulated more data (how much is unclear) on observer effects in forensic science in general and DNA profiling in particular. Ostrum remarks that sequential unmasking "may well result in new and unanticipated issues that end up being worse than the original concern." This worry is vague, particularly given the fact that sequential unmasking has already been instituted in some crime labs. Indeed, sequential unmasking is nothing more than the application of common practice in other sciences to forensic science.

It would be wrong to delay response to the possibility of observer effects until they have been "proven" to exist for experienced, fully trained examiners. First, the ubiquity of observer effects is "one of the better demonstrated findings of twentieth-century psychology" (12). Second, virtually no cost is associated with the implementation of sequential unmasking. Third, if a verdict of guilty requires proof beyond a reasonable doubt, then those who

deny the importance of observer effects in practice should bear the burden of proving their absence. Ostrum concedes all that is required when he states, "these issues may ultimately prove to be important factors in forensic decision-making." When in doubt, sequentially unmask.

## References

- Ostrum B. Commentary on: sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. *J Forensic Sci* 2009;54(6).
- Wells JD. Commentary on: sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. *J Forensic Sci* 2009;54(2):500.
- National Research Council. Strengthening forensic science in the United States: a path forward. Washington, DC: National Academy Press, 2009.
- Miller LS. Bias among professional document examiners: a need for procedural change. *J Police Sci Adm* 1984;12:407-11.
- Risinger DM. Appendix to goodbye to all that, or a fool's errand, by one of the fools: how I stopped worrying about court responses to handwriting identification (and forensic science in general) and learned to love misinterpretations of *Kumho Tire v. Carmichael*. *Tulsa L Rev* 2007;43:477-596. 481-82.
- Miller LS. Procedural bias in forensic examination of hair. *Law Hum Behav* 1987;11(2):157-63.
- Dror IE, Charlton D. Why experts make errors. *J Forensic Ident* 2006;56(4):600-16.
- Dror IE, Charlton D, Peron A. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci Int* 2006; 156:4-8.
- Wertheim K, Langenburg G, Moenssens A. Report of latent print examiner accuracy during comparison training exercises. *J Forensic Ident* 2006;56(1):55-127.
- Perrow C. *Normal accidents: living with high-risk technologies*. Princeton, NJ: Princeton University Press, 1984.
- Guerin B. What do people think about the risks of driving? Implications for traffic safety interventions. *J Appl Soc Psychol* 1994;24:994-1021.
- Nisbett R, Ross L. *Human inference*. Englewood Cliffs, NJ: Prentice Hall, 1980.

Dan E. Krane,<sup>1</sup> Ph.D.; Simon Ford,<sup>2</sup> Ph.D.; Jason R. Gilder,<sup>3</sup> Ph.D.; Keith Inman,<sup>4</sup> M.Crim.; Allan Jamieson,<sup>5</sup> Ph.D.; Roger Koppl,<sup>6</sup> Ph.D.; Irving L. Kornfield,<sup>7</sup> Ph.D.; D. Michael Risinger,<sup>8</sup> J.D.; Norah Rudin,<sup>9</sup> Ph.D.; William C. Thompson,<sup>10</sup> J.D., Ph.D.; and Marc S. Taylor,<sup>11</sup> B.S.

<sup>1</sup>Wright State University—Biological Sciences, 3640 Colonel Glenn Highway, Dayton, OH 45435-0001.

<sup>2</sup>Lexigen Science and Law Consultants, Inc., San Francisco, CA.

<sup>3</sup>Forensic Bioinformatics, Inc., Fairborn, OH.

<sup>4</sup>Forensic Analytical Sciences, Inc., Hayward, CA.

<sup>5</sup>The Forensic Institute, Glasgow, UK.

<sup>6</sup>Fairleigh Dickinson University, Institute for Forensic Science Administration, Madison, NJ.

<sup>7</sup>University of Maine, Molecular Forensics Laboratory, Orono, ME.

<sup>8</sup>Seton Hall University School of Law, Newark, NJ.

<sup>9</sup>Forensic DNA Consultant, Mountain View, CA.

<sup>10</sup>Department of Criminology, Law and Society, University of California, Irvine, CA.

<sup>11</sup>Technical Associates, Inc., Ventura, CA.