TECHNICAL NOTE

Yaron Shor,[1] *M.Sc. and Sarena Weisner,*[1] *B.Sc.*

A Survey on the Conclusions Drawn on the Same Footwear Marks Obtained in Actual Cases by Several Experts Throughout the World

REFERENCE: Shor Y, Weisner S. A survey on the conclusions drawn on the same footwear marks obtained in actual cases by several experts throughout the world. J Forensic Sci 1999;44(2): 380–384.

ABSTRACT: A survey was conducted with two sets of shoeprints from actual crime scenes and the corresponding suspect's shoes. Experts from seven countries were asked to give their opinion on the probability that the suspect's shoe made the shoeprint impression at the crime scene.

Each expert gave his/her opinion based on the scale used in his country. The distribution of the answers is discussed in this paper.

KEYWORDS: forensic science, criminalistics, shoeprints, footwear imprints, survey, identification scale

The results of the present survey are based on two actual cases sent to some forensic science laboratories. The survey concerned conclusions regarding the same shoeprint cases and are drawn by different experts.

Two methods for running such a survey are mentioned in the literature:

(1) A test is done with a known shoe and the expert usually receives impressions of the suspect's shoe and photographs of footwear impressions. There is a "correct" answer known to the distributor of the test. Such is the "Proficiency Test," done by Collaborative Testing Services, Inc. (1) and answered by many laboratories [about 180 labs in 1996 report (no. 9612)].

(2) A shoe from a known origin is selected, with authentic characteristics, and some "questionable" impressions are made. The distributor of the test determines the number of characteristics and marks them before sending the "known" prints with the "questionable" print for examination by the experts. The NBI lab in Finland submitted such tests and presented the results in the first European Meeting for Shoeprints/Toolmarks Examiners (2), and in the second European Meeting (3).

In this paper we propose a third way, which in our opinion is more realistic, and we applied it in our survey:

(3) Real cases were used. Test impressions of the suspect's shoes were prepared and then the test impressions and the photographs of the shoeprints from the scene of the crime were sent to each individual laboratory that was taking part is the test. In this kind of test, neither the maker of the test nor the participants know the "right" answer, as in real-life ambiguous cases.

Experimental

Two actual cases were chosen for the comparison. Both of them are ambiguous and controversial due to the nature of the vague imprint at the crime scene and the difficulties in finding the obscure individual characteristics (if they exist at all). The documentation in these cases was only a photograph that was taken by a technician when he arrived at the scene of crime Figs. 1, 2. The test impressions were made according to the normal procedure in our laboratory. We dusted the suspect's shoe sole with a mixture of gray and black fingerprint powder, and a person with the same shoe size as the suspect shoe wore the shoe and stepped on adhesive tape. The tape was then covered with clear celluloid (Figs. 1, 2) according to one of the procedures described by Bodziak (4).

These particular cases were chosen because when first examining them, the opinions of the different Israeli experts differed greatly and much time was spent analyzing them until a common conclusion was reached. Although a uniform conclusion was ultimately reached, we wanted to know if our conclusion would be a common opinion among other shoeprint experts throughout the world.

All the participating laboratories were given photographs of shoeprints from the scenes of crime in 1:1 scale, with 1:1 photographs of the suspects' shoes and actual test impressions for each case. The material was sent to ten laboratories in nine countries. The laboratories chosen are among the leading shoeprint laboratories in Europe and the United States. All the experts who took part in the test were qualified and experienced experts. Seven laboratories in six countries replied, and we received answers from 20 experts. The opinions given by three experts from Israel are presented as well.

The participants were asked to give their expert opinion about the possible connection between the shoeprint that was found at the scene of crime and the suspect's shoe. The answers were to be given in their own words according to the terminology used in their country. As in shoeprint expert opinions (in Israel), they were not asked to specify the number of individual characteristics they

[1] Scientific officers, Toolmarks and Materials Laboratory and Fingerprints Developing Laboratory, respectively, Division of Identification and Forensic Science, Israel National Police Headquarters, Jerusalem, Israel.

Received 11 Nov. 1997; and in revised form 6 April 1998; accepted 14 July 1998.

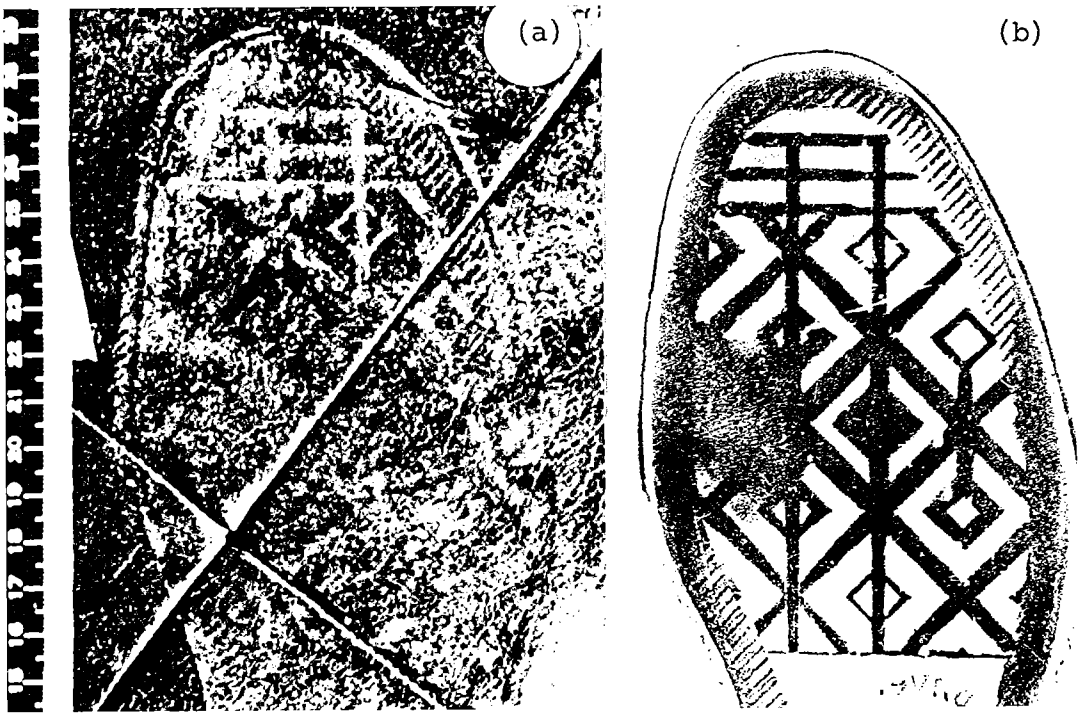

FIG. 1—*(a) Shoeprint photographed at the scene of crime. (b) Test impression from the suspect's shoe ("C.A.S").*

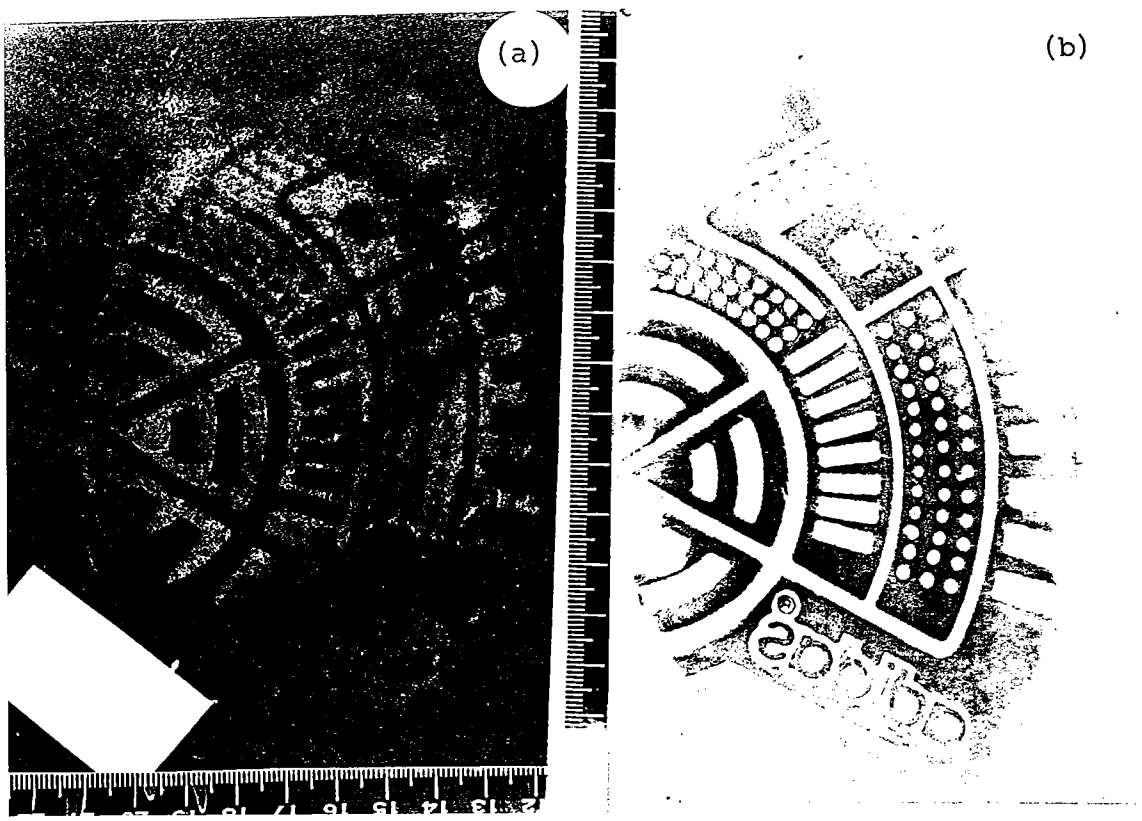

FIG. 2—*(a) Shoeprint photographed at the scene of crime. (b) Test impression from the suspect's shoe ("Adidas").*

had found. Even so, several laboratories did send the worksheets with the points that were used in determining the conclusion. Unlike the surveys conducted by Majamaa et al. (2,3), the manager of the survey had not indicated any characteristics to be considered.

In one of the cases a "Converse All Star" (C.A.S) shoe was involved (Fig. 1), and the second case involved an "Adidas" shoe (Fig. 2).

The wording of each expert reflects the scale that is used in his/her country. On obtaining all the replies, we transferred all the answers to the scale that is used in Israel (5).

Results and Discussion

In the textbooks by Bodziak (4) and Cassidy (6) as well as Abbott and Germann (7) the range of inconclusive conclusions is given briefly. The term "inconclusive" includes everything less than a positive identification. On the other hand, in Europe and several states within the United States (Hamm E. Florida Department of Law Enforcement, personal communication) the "Inconclusive" range is given in fine-tuned details. In published surveys (2,3,8) and presentations (5), the conclusions "inconclusive," "possible," "probable," and "highly probable" are described in detail.

In our work (8) we published a graphic figure that we called "The Identification Triangle Scale." The triangle is divided into six layers that represent levels of identification from "negative" to "full identification." Each term used in the triangle scale represents a range of certainties. The different heights of the levels of identification on the triangle scale illustrate this. It can easily be seen that the term "possible" represents a wide range of meanings, while the term "identification" has a very precise meaning. Two "possible" answers may be far from getting the same weight: one may be drawn near the "inconclusive" line, and the other may be situated near the "probable" line, while "identification" cannot be understood in more than one way. The numbers written near each category on the identification triangle scale represent the number of expert opinions given by all the participants in the survey (Figs. 3 and 5).

The second tool with which we used to illustrate the differences in the expert opinions is called "the Diversion Graph." The diversion graph was used to present the range within each laboratory. The *X*-axis represents the number of the laboratory. The heights between the lines on the *Y*-axis resemble the different distances in the "Identification Scale." In this scale it is also well shown

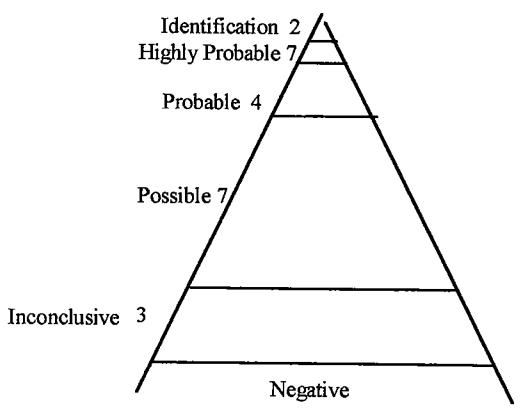

FIG. 3—*Identification triangle scale for the "Converse All-Star" shoe. The numbers indicates the number of experts in each category, total: 23 experts.*

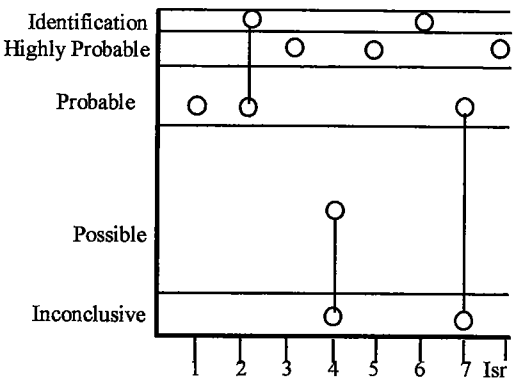

FIG. 4—*Diversion of opinions graph for the "Converse All-Star" shoe.*

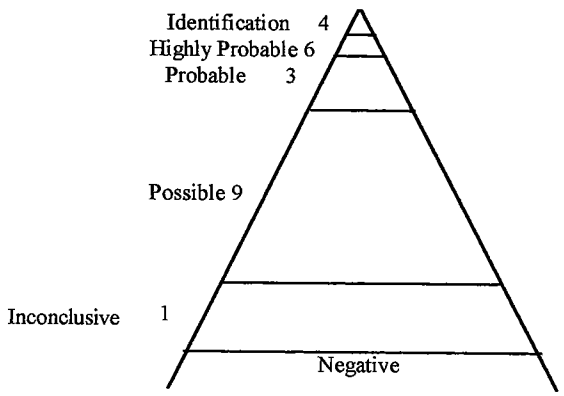

FIG. 5—*Identification triangle scale for the "Adidas" shoe. Numbers indicate the number of experts in each category, total: 23 experts.*

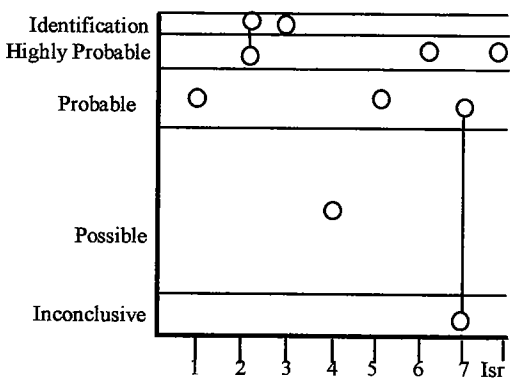

FIG. 6—*Diversion of opinions graph for the "Adidas" shoe.*

that even in the same laboratory, the diversion can be quite extensive. (This is illustrated by a vertical line, linking all the laboratory's expert opinions (Figs. 4,6).

All the answers were collected and translated to the wording with the closest meaning that appears in the survey done by Majamaa et al. (2,3). The terms "cannot be ruled out" and "cannot be eliminated" (Lab. No. 7) were translated to "inconclusive," "could have" and "could well have been made" was transferred to "possible" and "probable" accordingly.

Twenty-three experts returned their answers. The conclusions were given in the usual procedure of each participating laboratory,

TABLE 1—*Distribution of conclusions for C.A.S. shoes case.*

No. of Laboratory	No. of Experts	Exact Wording
1	1	probably caused by that shoe
2	1	full identification
	2	highly probable
	2	probable
3	1	highly probable
4	3	possible
	2	inconclusive
5	1	very probable
6	1	identification
7	1	can't be ruled out
	4	could have made
	1	could well have made
Israel	3	highly probable

TABLE 2—*Distribution of conclusions for Adidas shoes case.*

No. of Laboratory	No. of Experts	Exact Wording
1	1	probable
2	3	full identification
	2	highly probable
3	1	positively identified
4	5	possible
5	1	probable
6	1	very probable
7	1	cannot be eliminated
	4	could have made
	1	could well have made
Israel	3	highly probable

and they varied from "inconclusive" to "full identification" (Tables 1 and 2).

It can be seen that almost half (10 of 23) of the responses were in the lower range ("inconclusive" and "possible"), and the rest were given in the upper part of the scale ("probable," "highly probable" and "identification").

Shoeprint comparison is a branch of forensic science. One of the principles of the scientific method is the consistency of results when an experiment is repeated, regardless of time or place. This principle lies at the basis of proficiency tests—there is a "right" answer that every expert should reach, and a "wrong" answer. In the 1996 proficiency test (1) (report no. 9612) 93% of the participants answered correctly and in 1995 (report no. 9512) 95% gave the "right" answer (1). This trend was also observed by Peterson et al. (9), who calculated an average of 87% "right" answers in the years 1985 to 1991.

The proficiency test results are homogenized due to two factors: First, the proficiency test has a "right" answer known to the writers of the test, and usually the test chosen is a clear-cut one, and not like the questionable shoeprints sent on this survey. Second, there are only three acceptable answers to the proficiency test, "Yes," "No" and "Inconclusive." These two factors narrow down the answers given to the proficiency test, and reduce the variability of the answers.

On the other hand, our survey allowed every expert to use the terminology and scale used in his country. This approach was held by Majamaa and Ytti (2), and Majamaa, Virtanen and Ytti (3) in the surveys they conducted as well. Their conclusion was that different laboratories reached considerably different conclusions regarding identical hypothetical cases. The results of this survey, performed with true cases, indicated the same trend and the answers received were far from alike. Interestingly, the answers varied not only between individual experts, but also between countries. Laboratories No. 2 and 3, for example (Figs. 4 and 6), consistently gave highly conclusive results, while Laboratories No. 4 and 7 consistently reached a lower level of identification.

A simple explanation for this phenomenon is that laboratories No. 4 and 7 are not as experienced or trained as the other laboratories and this is what led to their answers.

Our opinion is quite different. Even regarding the same characteristics, answers given by experts from different countries (and probably different disciplines), may vary greatly. Evaluating and weighting all the factors involved in shoeprint examination may cause different experts to reach completely different conclusions based on the same facts. This happens often because of the lack of common standards in shoeprint examination. The number of different individual characteristics is almost infinite, unlike the final number of categories in fingerprints or DNA. For example, the obscureness of the photograph received may cause the expert to reduce the level of identification to a lower level than he might have reached if the details on the photograph were more noticeable. The same is true concerning the degree of wear or even the general shape of the ambiguous individual characteristics.

For example, the general shape of the ambiguous individual characteristics or even the degree of wear may cause the expert to reduce the level of identification to a lower level than he might have reached if the details on the photograph were more noticeable. The same is true concerning the obscureness of the photograph received. All of these factors may cause one expert to decide "cannot be eliminated," while his colleague, as trained and skilled, but guided by different assumptions, will reach a "high probability" conclusion.

This might agitate the shoeprints examiners' community, because one expert's opinion at court may be a "full identification," while his colleague (from another country) will present a mere "cannot be ruled out" opinion.

One might think that this diversion occurs only in those areas of forensic science where the comparative parameters are not solid, like a shoeprint, but this is true also for fingerprint identification. Fingerprint comparison is based apparently on rigid standards, as to what a point of comparison is, and in many countries there is even a set number of points needed in order to claim full identification. Nevertheless, as Evett et al. (10) have shown, set standards were not enough to prevent the participants of the survey from reaching conclusions with a broad scattering.

It is the authors' view that this variance cannot be eliminated, but it can be narrowed by proposing a set of guidelines that would assist the expert, guide him on his path, and lead him to a "right" conclusion. As Rudram concluded in his survey (11): "Forensic scientists need to agree amongst themselves what they mean by the probabilistic phrases they use." Our survey dealt utterly with shoeprints, but the results indicate the same trend of lack of clarity with the interpretation of forensic results.

Taroni and Aitken (12) claim that the likelihood ratio and the Bayesian framework are the only methods of evaluating forensic evidence in court. They propose a system based on the laissez-faire approach in which the numerical value of the likelihood ratio is mapped onto a verbal scale. Davis et al. (13) think that Taroni and Aitken's approach, though theoretically true, has many practical problems. They believe that verbal conclusions such as those presented by Rudram and in this survey are "a reasonable way *at present* of expressing the outcome of our work in terms that courts

will understand.'' In our opinion, setting an international terminology for the probability levels, and a set of guidelines for moving between the probability levels, will narrow the variance of answers given in different countries, and by different experts.

Acknowledgments

We wish to thank all the laboratories and the experts who took part in our survey for their patience and cooperation.

References

1. Collaborative Testing Services, Inc. Forensic Laboratory Proficiency Testing Program.
2. Majamaa H, Ytti A. A survey of the conclusion of similar footwear cases in various forensic laboratories. Forensic Sci Int 1996;Sept. 82(1):109–20. [Presented at the First European Meeting for SP/TM examiners in Helsinki, 1995].
3. Majamaa H, Virtanen J, Ytti A. Survey of the conclusion drawn of similar footwear cases, part 2, presented at the Second European Meeting for SP/TM examiners, The Netherlands, 1997. Proceedings in press.
4. Bodziak WJ. Footwear impression evidence. New York: Elsevier Science Publishing Inc. 1990.
5. Shor Y. The identification scale, presented at the Second European Meeting for SP/TM Examiners, The Netherlands, 1997.
6. Cassidy MJ. Footwear identification. Ottawa: RCMP GRC, 1980.
7. Abbott JR. Footwear evidence. Springfield: Charles C Thomas, 1964.
8. Shor Y, Even H. Scaling the term ''possible'' in shoeprints, (in Hebrew). Israeli J of Crim Just (Published by the Institute of Criminology and Criminal Law, Faculty of Law, the Tel-Aviv University), 1992;3:267–77 (English abstract).
9. Peterson JL, Markham PN. Crime laboratory proficiency testing results, 1978–1991, II: Resolving questions of common origin. J Forensic Sci 1995;Nov. 40(6):1009–29.
10. Evett IW, Williams RL. A review of the sixteen points fingerprint standard in England and Wales, J Forensic Ident 1996;46(1):49–73.
11. Rudram DA. Interpretation of scientific evidence. Sci Just 1996; 36(3):133–8.
12. Taroni F, Aitken C. Correspondence: interpretation of scientific evidence. Sci Just 1996;36(4):290:2.
13. Davis R, Facey O, Hamer P, Rudram D. Correspondence: interpretation of scientific evidence. Sci Just 1997;37(1):64–5.

Additional information and reprints request:
Yaron Shor, M.Sc.
Division of Identification and Forensic Science
Israel Police Headquarters
Jerusalem 91906, Israel