Gertrud A. A. Schoon,[1] *Ph.D.*

A First Assessment of the Reliability of an Improved Scent Identification Line-up

REFERENCE: Schoon GAA. A first assessment of the reliability of an improved scent identification line-up. J Forensic Sci 1998; 43(1):70–75.

ABSTRACT: To properly evaluate different forensic techniques, it is important to know how reliable these different techniques are. The reliability of scent identification line-ups is unknown. The purpose of this study was to describe, and employ, a reliability testing method for scent identifications using trained police dogs and a novel scent identification procedure. Two kinds of experiments were prepared: suspect = perpetrator experiments, and suspect ≷ perpetrator experiments. Six dog/handler teams participated in 10 experiments, five of each kind. The reliability of an identification, or the diagnostic ratio, is the percentage correct identification in suspect = perpetrator experiments divided by the percentage false identification of the suspect in suspect ≷ perpetrator experiments. Factors that influence the reliability of scent identifications are discussed, and the results of the scent identifications are compared with recent reliability estimates of other forensic techniques.

KEYWORDS: forensic science, scent, identification, line-up, canine reliability

The reliability of different fact-finding methods is a subject under debate. This debate has extended from "soft" psychological methods to "hard" scientific methods. The recent review of crime laboratory proficiency testing results by Peterson & Markham (1) illustrated that this debate is justified. Scientific methods may, in themselves, be reliable but it is the application in practice that yields the results presented as evidence in court. This process includes undetected technical and human errors from a variety of sources. Thus it is not sufficient to assess the reliability of evidence presented in court by looking at the theoretical reliability of the scientific method used, nor can one assess the reliability of the method by the level of training and practice of the scientist involved. The methods should be tested in full, and crime laboratory proficiency testing is an important first step.

This paper presents a reliability assessment of scent identification line-ups. In a scent identification line-up, a trained dog matches the odor of the perpetrator left on a corpus delicti to the odor of a suspect. The odor of the suspect is presented as one in an array of different odors, and the dog makes the match by only responding to the odor of the suspect and by ignoring the other odors. The two main assumptions underlying this method are that every human being has a (stable) unique odor, and that dogs are capable of discriminating between these odors. Although strictly speaking these assumptions have not been proven, a body of (circumstantial) evidence and empirical work supports them. The method is used as investigation tool by different police forces in the world, and is also presented as evidence in court (2).

Recent work has shown that the experimental set-up has a significant effect on the outcome of scent identifications. In a series of experiments using the same group of dogs, the way in which the odors in the array were presented and the "rules" surrounding the procedure were varied (3). Incorporating a "performance check" in the experimental set-up, where the dog's ability/willingness to work was tested directly prior to the scent identification, significantly enhanced the result of the identifications: there were both more correct identifications and less false responses. A second series of experiments with different dogs trained in a slightly different way confirmed these findings (2).

The "performance check" method was concluded to be a good experimental set-up for employing the capabilities of dogs for forensic purposes. To assess the reliability of any identification method, it is necessary to know the results in factually suspect = perpetrator cases as well as in factually suspect ≷ perpetrator cases. This paper describes the design and results of a study aimed at obtaining these data. The experiments mimicked forensic reality. The handlers knew that they were participating in "reliability" experiments but did not know more, which makes the results comparable to those obtained in proficiency testing of forensic detection methods where the forensic laboratories were also aware which material was part of the testing program.

Material and Methods

Animals

All experiments were done with six dogs trained and certified as Dutch "police human scent tracker dogs." Dogs 1–3 were male and dogs 4–6 were female. All dogs were Shepherd dogs (Malinois, German, Dutch, or mixed parentage) and their ages varied from 4–9 years. The dogs and their handlers worked in two groups: dogs 1, 2, 4 and 5 worked in area 1 and dogs 3 and 6 in area 2. The scent identification module of the compulsory yearly examination of these dogs differs in set-up from the one used in the experiments described here and has been described elsewhere (3).

Experiments

Each dog was used for 10 experiments: five "suspect = perpetrator" experiments and five "suspect ≷ perpetrator" experiments, using 5 different kinds of corpora delicti in both series. In an effort to minimize variables, one person was used as perpetrator and suspect in all "suspect = perpetrator" experiments, and two other

[1]University of Leiden, Department of Criminalistics and Forensic Science, Institute of Evolutionary and Ecological Sciences, Ethology group, National Dutch Police Tracker Dog Center, the Netherlands.

Received 29 Oct. 1996; and in revised form 9 April, 27 June 1997; accepted 27 June 1997.

people were used in all the “suspect ≳ perpetrator” experiments. The handlers were aware that an experiment could be of either type but at the time of the experiment they did not know what kind of experiment they were participating in. The dogs performed 2 experiments (one of each kind) on a test-day, test-days were 1–2 weeks apart.

Preparation of the Experimental Corpora Delicti

Two male civilian police employees volunteered as “perpetrators” for the experiments. They were asked to prepare corpora delicti in realistic ways. The corpora delicti used in the experiments were pistol buttplates, screwdrivers, spanners, sweatshirt cuffs, and scent samples taken from the seats of their car. The buttplates, screwdrivers and spanners were kept in the pocket for approximately 15 min (simulating a perpetrator carrying this material to the crime scene), and then handled for 5 min longer (simulating working with this equipment). The sweatshirt cuffs were worn around the wrist for 15 min and also handled for 5 min. The scent sample was taken according to standard police protocol by placing an odor collection cloth (cotton bandage) for 1,5 hours on the seat of the car of the “perpetrator.” The “perpetrator” had driven the car for 30–45 min on his way to work, and the scent sample was taken less than an hour after arrival. The corpora delicti were stored for 8–10 days according to customary police protocol in plastic bags or glass jars with a twist-off top.

Preparation of the Experimental Odor Arrays

In forensic reality, the suspect is usually the only person who is in jail and the other odors in the array belong to policemen or civilian police employees. To simulate this difference, the experimental “suspects” were from a different environment than the people who volunteered to prepare the decoy odors.

In the “suspect = perpetrator” experiments, the “perpetrator” who participated in preparing the corpora delicti was also the “suspect.” In the “suspect ≳ perpetrator” experiments, another male civilian police employee volunteered as “suspect.” The odor arrays were prepared following customary police protocol. Each of the “suspects” and the decoys (male and female police school students) were given 2 glass jars containing 6 stainless steel tubes each, and were asked to handle the tubes for 5 min per jar. The jars were marked to differentiate between the batch of tubes scented first and the batch scented second. Handling 12 tubes in all is more than usual in police practice, where handling 2–4 tubes in total is customary. In forensic practice the participants wash their hands prior to handling the tubes which was not done in these experiments. For each experiment, one of the male decoys was designated as “check” person for the performance check and this person was also requested to handle pieces of standard electric wire tubing (“PVC-tube”). The people who prepared the decoy odors only participated once in the series of experiments. The experiments were usually conducted the day after the experiments had been prepared.

Experimental Protocol

For each experiment, 14 tubes containing odors of 7 different people were arranged in two rows. Each row contained the odor of a “suspect,” the odor of a “check” person, and 5 other decoy odors. The position of these different odors was random with one limitation: in one row the odor of the “suspect” would come before the odor of the “check” person, in the other row this would be the reverse. This was done to maximize the chance of the dog smelling and ignoring the odor of the “suspect” during the performance check (this demonstrated the lack of specific interest of the dog for this particular “suspect”). The first row contained tubes from the jar that was prepared first, the second row contained tubes from the jar that was prepared second. The rows were prepared in absence of the dog handler, who did not know the position of the matching tubes at any time during an experiment.

In the first two trials the dog was given a “performance check.” Through a simple test the ability/willingness to work was established.

For trial 1, the dog was given the “PVC-tube” handled by the “check”-person as a sample odor, and had to find the matching odor in the first array of seven tubes. If the dog retrieved a non-matching tube or did not retrieve any tube at all, the result was “D(isqualified)1” and experiment was terminated. If the dog retrieved the matching tube, this was noted as “correct” and the dog and his handler continued.

In trial 2 the match was repeated in the second array of odors: the dog was given the same “PVC-tube” as a sample and had to retrieve the tube containing the odor of the “check” person in the second row. Failing to retrieve, or retrieving a non-matching odor, led to a “D(isqualification)2.” If the dog retrieved the matching tube again, this was noted as “correct” and the experiment continued.

After succeeding in trials 1 and 2, the dog was considered “qualified” to work. Both rows now contained 6 tubes with the odors of the suspect and 5 decoys, since the tubes containing the “check” odor had been removed by the dog.

For trial 3, the dog and the handler returned to the first array. The dog was given the odor of the perpetrator on a corpus delicti as a sample odor, and has to find a matching tube. If the dog retrieved the tube of the suspect, the experiment continued with trial 4. Not retrieving anything was noted as “0” (see Table 1: correct in “suspect ≳ perpetrator” cases, a miss in “suspect = perpetrator” cases). Retrieving a non-matching tube was noted as “ – ” (Table 1).

In trial 4, the dog had to match the odor of the perpetrator to that of the suspect a second time, but now in the second array of odors. Retrieving the tube containing the odor of the suspect a second time was noted as “ + ” (Table 1: correct identification in “suspect = perpetrator” experiments, and false identification in “suspect ≳ perpetrator” experiments). Retrieving a non-matching tube would have led to a “ – ” and not retrieving any tube to “0” but these situations did not arise.

A flow chart and a schematic overview of the possible results of each experiment is given in Table 2. The trials were videotaped for further analysis.

Results

The results of the 10 experiments per dog are given in Table 3.

Half of the experiments led to a disqualification in the first or second trial. As can be seen in Table 3, 19 of the 30 “suspect = perpetrator” cases ended in a D1 or D2, as well as 11 of the 30 “suspect ≳ perpetrator” cases. Analyzing these disqualifications further shows that 21 disqualifications were a result of a mistake in trial 1, and 9 the result of a mistake in trial 2. A division could be made into a group with disqualifications predominantly in trial 1 (dogs 1, 2 and 4) and a group with 50% or more disqualifications in trial 2 (dogs 3, 5 and 6). Trial 2 consisted of tubes that were handled second and could, therefore, contain less odor, so the behavior of these two groups was analyzed further. After a correct

TABLE 1—*Registration of the results of the choices made by the dogs.*

Reality ⇓	Choice of Dog: Odor Suspect	Odor Decoy	No Odor
suspect = perpetrator	+: correct identification	−: wrong	0: miss
suspect $\gtrsim$ perpetrator	+: false identification	−: wrong	0: correct non-identification

retrieval in trial 1, dogs 3, 5 and 6 are disqualified more often in trial 2 (28%) than the other three dogs (13%). When dogs 3, 5 and 6 do succeed in both trials 1 and 2, they retrieve the correct tube more quickly in trial 1 than in trial 2 in 12/16 experiments. The other three dogs retrieve the correct tube more quickly in trial 2 than in trial 1 in 8/10 experiments, which is what one would expect since trial 2 is a simple repetition.

A second observation regarding the disqualifications is that dog 4, who was disqualified in all of the "suspect = perpetrator" experiments, showed significant interest for the odor of the "suspect" in trials 1 and/or 2 of these experiments. The odor of the "check" person was different in each of these experiment. In two experiments dog 4 retrieved the "suspects" tube in trial 1 (instead of the "check"-tube), in the other 3 she demonstrated such interest in the "suspects" tube that the handler thought that this was the correct "check"-tube. Dog 4 only showed this interest for the "suspect" in the "suspect = perpetrator" experiments and not in the "suspect $\gtrsim$ perpetrator" cases.

A third observation regarding the results is the difference in mental states of the handlers in the two groups. In area 1 (dogs 1, 2, 4 and 5) the majority of the handlers said that they were very nervous about the experiments. In area 2 (dogs 3 and 6) the handlers were more relaxed and confident. The experiments in area 1 led to significantly more disqualifications than in area 2 (60 vs. 30%, χ^2, $p = .03$).

In Table 4, an overview of the total results of the experiments is given. In the "suspect = perpetrator" experiments the suspect was correctly identified as the perpetrator in 4 of the 11 experiments where the dogs were qualified. They made a choice in 9 of these experiments, 4 of which were correct choices. Since they have a choice out of 6, this is better than chance ($p < .05$, binomial test).

In the "suspect $\gtrsim$ perpetrator" tests 9 of the 19 experiments where the dogs were qualified led to the correct response: no retrieval at all. In the other 10 experiments the dogs did retrieve a tube, and this led to a false identification of our "suspect" once. Since they have a choice out of 6, this is not significantly different from chance.

Discussion

In forensic investigations in general, the "reality" as described in Table 1 is not known. One only knows if the result of an investigation is "positive identification/same origin" or "non-identification/different origin." In order to translate experimental material such as the material collected in this study to a practical assessment of reliability, the "diagnostic ratio" (4) was calculated. This diagnostic provides insight into how often a method is correct

TABLE 2—*Flow chart of an experiment. Per trial, each possible response a dog may give is listed (retrieval of whose odor, or no retrieval at all), followed by its consequences (end or continue, symbolized by ✣). In the last column the way the result of a total experiment is scored is given.*

Trial 1	Trial 2	Trial 3	Trial 4	Result
Decoy	End			D1
None	End			D1
Control	✣ Decoy	End		D2
	✣ None	End		D2
	✣ Control	✣ Decoy	End	—
		✣ None	End	0
		✣ Suspect	✣ Decoy	—
			✣ None	0
			✣ Suspect	+

TABLE 4—*Summarized results of the 60 experiments. The "correct" experiments are given in italics.*

	Disqualified	Qualified, Dogs Subsequently Retrieve Odor: Suspect	Decoy	None
Suspect = Perpetrator	19	*4 (36.5%)*	5 (45.5%)	2 (18.2%)
Suspect $\gtrsim$ Perpetrator	11	1 (5.3%)	9 (47.4%)	*9 (47.4%)*

TABLE 3—*Results of the 10 experiments per dog.*

	Suspect = Perpetrator					Suspect $\gtrsim$ Perpetrator				
	Buttplates	Screwdriver	Spanners	Shirtcuffs	Scent Sample	Buttplates	Screwdriver	Spanners	Shirtcuffs	Scent Sample
Dog 1	D1	+	+	+	D1	—	—	—	0	D1
Dog 2	D1	D1	—	D1	D1	—	—	0	D1	D1
Dog 3	+	0	D2	D2	D1	0	D2	0	0	—
Dog 4	D1	D2	D1	D1	D1	0	D2	D1	D1	D1
Dog 5	D2	D2	0	D1	—	—	D2	—	D1	D1
Dog 6	D1	D2	—	—	—	0	0	+	—	0

D1: disqualified in trial 1, D2: disqualified in trial 2.
+: positive identification of suspect as perpetrator after qualification.
—: wrong, retrieval of tube with decoy odor after qualification.
0: no identification, no retrieval of any tube after qualification.

when the result is "positive identification," and how often it is correct when the result is "negative identification." The ratio is calculated as follows:

$$\text{diag. ratio of "positive id."} = \frac{\text{\% correct id. in suspect} = \text{perpetrator cases}}{\text{\% false id. in suspect} \gtrsim \text{perpetrator cases}}$$

and

$$\text{diag. ratio of "negative id."} = \frac{\text{\% correct non-id. in suspect} \gtrsim \text{perpetrator cases}}{\text{\% misses in suspect} = \text{perpetrator cases}}$$

A diagnostic ratio of 10 means that for every 10 times a result is correct, it is incorrect once, whereas a diagnostic ratio of 20 means that for every 20 times a result is correct, it is incorrect once. At first sight, a method with 100% correct identification but with 20% false identification might seem better than one with a 20% correct identification and 2% false identifications. However, the first method leads to an incorrect result once for every 5 times it is correct, and the second to an incorrect for every 10 times it is correct. For an evaluation of the method, the diagnostic ratio is essential.

In an earlier study where different experimental set-ups were compared, only suspect = perpetrator cases were offered to the dogs (3). But by extrapolating the ratio between "wrong" and "miss" obtained in these experiments to suspect ≳ perpetrator cases, one can calculate a % false identifications (by dividing the total % wrong by the number of odors in the set-up) and % correct non-identifications. Thus applied for the above study, the diagnostic ratios for a "positive identification" was almost twice as high for the "positive check" method in comparison with the currently used experimental set-up.

The data obtained in this study (Table 4) can be used directly in the formula's, leading to diagnostic ratio's of 6.9 for a "positive identification" (36.4% correct identifications/5.3% false identifications) and 2.6 for a "negative identification" (47.3 correct non-identifications/18.2% misses). However, the data obtained in these experiments seem to be negatively biased for a number of reasons. After examining these reasons, an expected realistic value of the different percentages will be given, leading to a new estimation of realistic diagnostic ratio's.

Three dogs showed a high proportion of disqualification in trial 2, and a slowness to retrieve in this second trial. The tubes in this trial belonged to the batch that was handled second, and it seems possible that these findings were caused by the large amount of material that was scented by each person. The difficulties these dogs encountered in this second trial may have had a negative effect on their performance in trial 3. The nervousness of part of the handlers involved in the experiment could also have negatively biased the results. Prior work in "suspect = perpetrator" scenarios confirm that the percentage correct identifications is usually substantially higher (3 and 5). A realistic level is estimated at 60%.

The percentage false identifications in "suspect ≳ perpetrator" cases seems, at first sight, to be underestimated in this study: the dogs only performed 47% of these experiments correctly. They picked up one of the six tubes in 53% of the cases, which would lead to a chance false identification of 8.8% in trial 3. Earlier work has shown that not all mistakes are repeated in a subsequent trial: after an incorrect choice the dogs only repeated this in 36% of the following choices (2). Assuming a repeat-percentage of 50%, the level of false identifications would drop to 4.4% in trial 4 of the "suspect ≳ perpetrator" cases. This then leads to a new, estimated diagnostic value of 13.6 for positive identifications.

The data on the negative identifications seems in line with prior studies (3): 50% correct non-identification in "suspect ≳ perpetrator" cases, and 20% misses in "suspect = perpetrator" cases.

This means that when using the improved "positive check" method described in this paper, it is expected that there will be one falsely accused in every 13–14 "positive identifications," and one who is falsely acquitted in every 2–3 "negative identifications." Therefore, "positive identifications" are more reliable than "negative identifications." This asymmetry is common in forensic science but one that the judicial system should be aware of.

Comparing the ratio on the "positive identification" with "positive on common origin" ratio's calculated from Peterson & Markham's overview (1), scent identifications can be placed in the "moderate success" group together with bloodstain analysis, questioned documents, toolmarks and hair analyses (diagnostic ratios: 10.0–29.4). Analyses on paint, glass, fibers and body fluid mixtures are less reliable and described as a category of concern (diagnostic ratios 3.1–7.8); fingerprints, firearms and footwear analyses were performed best (diagnostic ratios 52.9–160.8). Experiments with eyewitness confrontations have led to diagnostic values varying between 9 (6) and 15 (7).

The experimental set-up used in these experiments differs from the one the dogs are certified in. In the "certification set-up" the odor of the suspect is one in an array of 12 different odors on tubes, which are divided into two rows of six tubes each. The dogs thus have to find the matching odor in one row, and refrain from making a match in the other row. The main differences between the "certification set-up" and the "performance check set-up" are: (a) the performance check on the dogs ability/willingness to work, (b) the check that the dog does not "prefer" the suspect, and (c) to obtain a negative identification a dog may not respond to any odor twice in the certification set-up, but only once in the performance check set-up.

A performance check prior to the actual forensic question was shown to have a significant positive effect on the results (3). The level of performance directly influences the reliability of the results obtained, which means that if the dogs are not able or willing to work, one should not use these dogs for forensic testing. Even the relatively simple check with a well-scented control object leads to a better performance. One explanation for this is that the simple performance check could be sufficient to test the olfactory ability of the dogs, which can vary due to hormonal changes, possible infections, (cross)adaptation or illness. This would mean that mistakes made by dogs that are not sufficiently able are eliminated by the performance check. Another explanations is that the simple check really tests willingness to work, thus eliminating mistakes made by unwilling dogs. Correct retrievals in the first two trials may even enhance the willingness to work by creating a "winning mood." Willingness to work is part of the very complex "motivation" of the dog, which is probably also influenced by the handler as demonstrated by the difference in results between the two groups. Which factor (olfactory ability or willingness, or both) is responsible is perhaps not directly relevant for the results but is of definite interest for training and selection of the dogs. If ability is the crucial factor, selecting dogs on olfactory ability could lead to long term improvement, if willingness is the crucial factor, more attention should be paid to enhancing the motivation of the dogs.

The second effect of the performance check is that one shows that the other odors in the array, including the odor of the suspect, are "neutral" for the dog and that there is no prior preference of the dog for the odor of the suspect. This neutrality is important in court: a positive identification may not be the result of a particular preference the dog may have, or because the odor of the suspect is very different from the others in the array. This has led to rules for the odors presented in the array in forensic tests. Since little is known about what determines scent, these rules have been based on what is known to influence comparable visual identifications, essentially minimizing differences between the people who participate in preparing the array. Thus the rules say that people have to belong to the same sex, must have the same racial background, and all must wash their hands with non-perfumed soap prior to scenting the tubes.

As we learn more of what determines human odor the list of rules will become longer. For example: since the influence of the major histocompatibility complex (a group of genes responsible for the human immune system) has been shown to have a significant effect on human odor (8,9) one might say that the odors in the array should belong to MHC compatible people. This is extremely costly (tissue-typing people is expensive) and it is practically impossible to find compatible MHC types, as is well illustrated by the compatibility difficulties in organ transplantations that are also a result of different MHC types. Another example: if a "stress-odor" exits, one might say that all the people in the array should be equally "stressed" as the suspect. However, rules will never suffice as long as it is unknown which components of human odor the dogs use for their discrimination.

Up to now it has not been shown that dogs use specific selection rules when responding. For example, it has been shown that dogs do not use olfactory information on sex or smoking habits to narrow down their choice (2). More importantly, the current rules do not guarantee that the dog did not respond to the odor of the suspect because he found it of interest. The "performance check set-up" tries to guarantee the "neutrality" of the row by letting the dog smell, and ignore, the other six odors in the first two trials of the set-up. The continued preference of dog 4 for the suspect in the "suspect = perpetrator" cases demonstrates the necessity of such a precaution. Since the "neutrality" of the odors is demonstrated for each particular dog at the time of the experiments, the necessity of rules surrounding the odors in the array is eliminated.

The third point of difference between the two set-ups concerns the "negative identifications." Experiments with mice and rats on odors in "go-no go" paradigms have shown that the majority of the mistakes are made by the animal responding in the "no go" situation, thus creating "false hits" (10). Not doing something was concluded to be more difficult than doing something, and this aspect is also a problem in scent identifications that follow a match-to-sample paradigm. When no match is possible, for example when a suspect is innocent (or for any other reason), the dogs should not respond at all. In the "certification set-up" a suspect $\gtrsim$ perpetrator situation means not responding twice in succession. In the "experimental set-up" a dog first has to perform the two performance check trials where it should respond (and be rewarded), after which it is only confronted once with an array where it may not respond. Thus the dog is confronted less with the difficult situation of not being allowed to respond in the "experimental set-up," and this was shown to have a positive effect on the reliability (3).

Not responding in suspect $\gtrsim$ perpetrator cases is a crucial aspect of reliability. It seems important to keep a constant check on the dogs for incorrect responses since this increases the chance of false identifications. The handlers should be made very aware of this problem since they can influence these mistakes. In normal training, the dogs are allowed two or three passes over the array of tubes to choose. If they do not choose, they are recalled by their handler. In forensic cases the handlers may be tempted to let the dog search for a longer time before recalling him: for example when the suspect has already confessed or when there are other compelling reasons to believe the suspect guilty. So although the handler is not aware of the position of the suspect's tube in the array, he may still influence the result. True blind experiments, where the handler does not know anything of the case at hand, is not a custom in police work but is probably the easiest way to prevent any (subconscious) influence. An alternative may be videotaping the forensic experiments for later examination. A third alternative is radically changing the experimental set-up into one where the dog can actively respond in both suspect = perpetrator and in suspect $\gtrsim$ perpetrator cases, as was successfully done in a pilot study with four dogs (11).

The conclusion can be drawn that scent identification following the improved "performance check" set-up described here are reliable enough to be a useful forensic tool. Odor is easily left behind by perpetrators and scent identifications can provide unique leads in forensic investigations. But: scent identifications should only be performed by dogs that are part of a comprehensive quality guarding scheme. In this scheme the performance level of the dogs should be monitored so that courts can be informed of the reliability of each particular dog if necessary.

Acknowledgments

First, I would like to thank the handlers who participated in these experiments. I would also like to thank Professor, Dr. P. Sevenster for his support during the experiments, and Professor, Dr. E. R. Groeneveld and Professor, Dr. C. J. ten Cate for their comments on the results and the manuscript.

References

1. Peterson JL, Markham PN. Crime laboratory proficiency testing results, 1978–1991, II: Resolving questions of common origin. J Forensic Sci 1995;40(6):1009–29.
2. Schoon GAA. The performance of dogs in identifying humans by scent [dissertation]. Leiden: University of Leiden, The Netherlands, 1997.
3. Schoon GAA. Scent identification lineups by dogs (Canis familiaris): experimental design and forensic application. Appl Anim Behav Sci 1996;49:257–67.
4. Malpass RS, Devine PG. Research on suggestion in lineups and photospreads. In: Wells GL, Loftus EF, editors. Eyewitness testimony: psychological perspectives. New York: Cambridge University Press, 1984.
5. Schoon GAA, Bruin JC de. The ability of dogs to recognize and cross-match human odors. Forensic Sci Int 1994;69:111–8.
6. Cutler BL, Penrod SD, Martens TK. The reliability of eyewitness identification: the role of system and estimator variables. Law Hum Behav 1987;11(3):233–58.
7. Wagenaar WA, Veefkind, N. Comparison of one-person and many-person lineups: a warning against unsafe practices. In: Lösel F, Bender D, Bliesenser TH, editors. Psychology and law: international perspectives. Berlin: Walter de Gruyter, 1992;275–85.
8. Wedekind C, Seebeck T, Bettens F, Paepke AJ. MHC-dependent mate preference in humans. Proc R Soc Lond B 1995;260:245–9.
9. Ferstl R, Eggert F, Westphal E, Zavazava N, Muller-Rucholtz W. MHC related odors in humans. In: Doty RL, Muller-Schwarze D,

editors. Chemical signals in vertebrates VI. New York: Plenum Press, 1992;205–11.
10. Lu XM, Slotnick BM, Silberberg AM. Odor matching and odor memory in the rat. Psychol Behav 1993;53:795–804.
11. Schoon GAA. Scent identification by dogs (Canis familiaris): a new experimental design. Behavior. In press.

Additional information and reprint requests:
Dr. G.A.A. Schoon, University of Leiden
Dept. of Criminalistics and Forensic Science
P.O. Box 9520
2300 RA Leiden
The Netherlands