SUPREME COURT COPY

N[redacted] S139103

IN THE SUPREME COURT OF CALIFORNIA

PEOPLE OF THE STATE OF CALIFORNIA,

Plaintiff and Respondent,

vs.

BAILEY JACKSON

Defendant and Appellant.

SUPREME COURT
FILED

JUN 27 2012

Frederick K. Ohlrich Clerk

Deputy

Automatic Appeal from the Superior Court
of Riverside County
Case No. RIF097839
Honorable Patrick F. Magers, Judge

APPENDIX TO APPELLANT'S OPENING BRIEF

RICHARD I. TARGOW
Attorney at Law (SBN 87045)
Post Office Box 1143
Sebastopol, California 95473
Telephone: (707) 829-5190

Attorney for Appellant

DEATH PENALTY

TABLE OF CONTENTS

APPENDIX A

J Forensic Sci, July 2006, Vol. 51, No. 4
doi:10.1111/j.1556-4029.2006.00178.x
Available online at: www.blackwell-synergy.com

Brian A. Eckenrode,[1] *Ph.D.; Scott A. Ramsey,*[2] *M.S.; Rex A. Stockham,*[3] *M.F.S.; Gary J. Van Berkel,*[4] *Ph.D.; Keiji G. Asano,*[4] *M.S.; and Dennis A. Wolf,*[5] *Ph.D.*

Performance Evaluation of the Scent Transfer Unit™ (STU-100) for Organic Compound Collection and Release

---

ABSTRACT: The Scent Transfer Unit™ (STU-100) is a portable vacuum that uses airflow through a sterile gauze pad to capture a volatiles profile over evidentiary items for subsequent canine presentation to assist law enforcement personnel. This device was evaluated to determine its ability to trap and release organic compounds at ambient temperature under controlled laboratory conditions. Gas chromatography-mass spectrometry (GC-MS) analyses using a five-component volatiles mixture in methanol injected directly into a capture pad indicated that compound release could be detected initially and 3 days after the time of collection. Additionally, 15 compounds of a 39-component toxic organic gaseous mixture (10–1000 parts per billion by volume [p.p.b.$_v$]) were trapped, released, and detected in the headspace of a volatiles capture pad after being exposed to this mixture using the STU-100 with analysis via GC-MS. Component release efficiencies at ambient temperature varied with the analyte; however, typical values of *c.* 10% were obtained. Desorption at elevated temperatures of reported human odor/scent chemicals and colognes trapped by the STU-100 pads was measured and indicated that the STU-100 has a significant trapping efficiency at ambient temperature. Multivariate statistical analysis of subsequent mass spectral patterns was also performed.

KEYWORDS: forensic science, volatiles profile, scent, odor, canine, scent transfer

The landmark Supreme Court case, *Daubert vs. Merrell Dow Pharmaceuticals* (1993) (1), significantly changed the legal admission procedures for scientific evidence in criminal cases. Previously, the 1923 decision in *Frye vs. United States* (2) stated that a procedure or method claiming to have scientific merit had to be considered reliable in the scientific community, have testimony provided by a qualified subject expert, and have proof presented that the person performing the test used correct scientific procedures. *Daubert* went further by adding specificity to the requirements that scientific claims must be verifiable, published in scientific journals, have a known error rate, and be standardized. As this decision, a number of common methods utilized by law enforcement and forensic science laboratories have undergone scrutiny in the United States judicial system. One of these methodologies involves scent-discriminating canines and the investigative tools utilized.

For admissibility purposes under *Frye* and *Daubert*, it would be beneficial for law enforcement purposes to reproduce the detection capability of canines in methodical form, with real-time and onsite capabilities. Research efforts are focusing on improvements in sampling of organic chemicals emanating from clothing or other evidentiary material, and furthering the development and use of effective trapping materials to characterize and facilitate the canine's scent-sensing ability. This research presents results from a performance evaluation of a device called the Scent Transfer Unit™ (STU-100), which is currently used by law enforcement for sampling the volatiles profile emanating from evidentiary material.

The history of using human scent-discriminating canines (*Canis familiaris*) for searching and for identifying criminal suspects in the United States is extensive. Positive scent matches are routinely accepted in the criminal justice system as probable cause and can be admitted as evidence in a trial, provided that additional evidence corroborates the canine's response (3). Unlike many European nations that have national certification and proficiency standards for scent-discriminating canines (4,5), there are no national standards in the United States. As a result, the reliability of scent-discriminating canines to correctly match and identify individuals from scent objects continues to be debated (6–11).

There are essentially six categories of detection canines being used by law enforcement: trailing, tracking, article detection, substance detection, area search, and scent identification line up. Trailing canines are trained to match the volatiles profile (scent/odor) acquired from an article of evidence to a matching trail of scent/odor present on the ground or in the field. Tracking canines are trained to follow ground disturbances, crushed vegetation, and although a human odor component may be present, they are not required to match a scent sample. Article detection canines are trained to locate items recently deposited within a search area. Substance detection canines used for detecting the presence of narcotics, explosives, arson accelerants, or human remains are typically trained on predetermined specific chemicals or mixtures and they are taught to alert when a match is located. Area search-and-rescue canines are trained to search mass disaster areas for the presence of live humans. Scent identification line-up canines are trained to use the scent/odor acquired from an article of evidence to identify the suspect of a crime from a line up of scented objects.

[1] Federal Bureau of Investigation, Counterterrorism/Forensic Science Research Unit, FBI Academy, Quantico, VA 22135.

[2] Federal Bureau of Investigation, Visiting Scientist Program through Oak Ridge Institute for Science and Education, FBI Academy, Quantico, VA 22135.

[3] Federal Bureau of Investigation, Explosives Unit, Quantico, VA 22135.

[4] Oak Ridge National Laboratory, Organic and Biological Mass Spectrometry Group, Oak Ridge, TN 37830.

[5] Oak Ridge National Laboratory, Computer Science and Mathematics Group, Oak Ridge, TN 37830.

Received 23 Oct. 2005; and in revised form 4 March 2006; accepted 26 March 2006; published 23 June 2006.

A human volatiles profile is more chemically complex and requires substantially different canine training scenarios compared with those used with targeted chemicals.

For the purposes of the canine and law enforcement communities, the terms scent and odor are community specific and are not necessarily interchangeable with other disciplines, such as biology, ecology, etc. Progress is being made in characterizing the variables and defining the terminology behind human scent composition and canine olfactory systems to develop a better understanding of the canine's response to mixtures of volatile and semivolatile chemicals. Curran et al. (12) developed terminology to categorize the complex mixtures that constitute human scent or odor. The first category, "primary odor," consists of constituents that are stable over time regardless of diet or environmental factors and are genetically based. A second category, "secondary odor," contains constituents that are present due to diet and environmental factors. Finally, "tertiary odor" contains constituents that are present due to the influence of outside sources such as lotions, soaps, or perfumes. For the purposes of human odor detection and analysis by canines, scent is considered the overall volatiles profile left by a human and odor consists of the elements of the volatiles profile that elicits a behavioral response. There may be dozens of compounds present in human scent, but it is unknown at this time the identity or quantity of compounds required by the canine to obtain an odor-related behavioral response for match-to-sample recognition. A volatiles profile is the terminology used here for describing a superset of both scent and odor that is independent of the human model (odor) and the canine model (scent/odor) as well as an instrumentation model (sampling chemicals) as it is known that particles are not responsible for eliciting a volatiles detection response in canines (13). The purpose of this study was to explore the ability of a specific sorbent to collect and release a combination of volatile chemicals. Although the compounds selected for this study have been previously reported to be components of human scent, the authors do not intend to imply that they are the components utilized by canines.

STU-100

There are four commonly used methods to collect a human volatiles profile for canine use: direct, swipe, adsorption/absorption, and indirect. The traditional direct method allows the canine to smell an article of evidence or volatiles source directly by bringing the item close to, or in contact with, its nose. Swiping involves wiping the surface of the evidentiary material with a sterile gauze pad and thereby transferring volatiles onto the pad. The pad is then presented to the canine as in the direct method. Adsorption/absorption involves placing a sterile gauze pad on the source surface for some time period, thereby creating a concentration effect or an aggregate from the item(s) of interest. Often, the source object(s) and a sterile gauze pad are placed into and sealed in a plastic resealable bag. After some time, the gauze pad is removed and presented to the canine as previously described (3). A major drawback to these three methods is the possible disruption and contamination of trace evidence *within* the object of evidence during scent pad contact. To address this issue, U.S. law enforcement personnel have recently been using an indirect or noncontact method of volatiles collection via the STU-100. Very little scientific data have been published that characterizes the trapping medium's (cotton pad) ability to adsorb and desorb organic compounds responsible for scent at ambient temperature.

Developed by Tolhurst and Harris and patented in 1998, (14) the STU-100 device is a portable, hand-held vacuum pump with a modified inlet, that is able to hold in place a 12.5 cm × 23.0 cm Johnson & Johnson® sterile surgical gauze pad. This "scent pad" is used as a trap to collect primarily volatile or vaporized scent compounds as the STU-100 pulls air through the gauze or sorbent at a flow rate of *c.* 300 L/min while it is physically swept above or over articles of evidence or over areas that may emanate volatiles. The pad is then removed from the STU-100 device and double packaged in heat-sealed nylon envelopes. To conduct a scent check with a trailing canine, the handler first acclimates the canine to the available volatiles profiles (scents and odors) at the start location and establishes a baseline for the canine. After harnessing, the handler opens the nylon envelope and places the pad in front of the canine's nose. If a matching odor is present at the trail start, the canine commences to follow the trail. If no matching odor is present, or the level of volatile organic compounds is below the detection capability of the canine, the canine is trained to respond by refusing to trail.

As the number of criminal investigations utilizing the STU-100 has increased, challenges to its court admissibility have surfaced. Officials of the Law Enforcement Bloodhound Association (LEBA) and the National Police Bloodhound Association (NPBA) have criticized the capabilities of the STU-100, although both organizations have members who own and use it. Published statements reflect that neither LEBA nor NPBA currently endorse the STU-100's "trap-and-release" capabilities (15). To date, neither organization has scientifically tested the device. The primary concern expressed by both bloodhound organizations and recent legal proceedings is that the STU-100 does not efficiently collect scent and that the device itself possibly contaminates the scent pads. A recent double-blind study using the STU-100 with canines from the Southern California Bloodhound Handlers Coalition showed that the percentage of positive-scent matches made by bloodhounds between human scent sampled from postblast debris and respective human subjects was 78.3%, with no false-positive identifications (3). For these tests, a false-positive identification is defined as a canine alert to a human subject whose odor was not present on the STU-100 pad during scent proffer. In another study using the STU-100, Harvey and Harvey (16) evaluated eight trained bloodhounds (three novices and five veterans) for the ability to discriminate scent between two human subjects and effectively trail the scent in a battery of terrain and weather conditions. (*Note*: Bloodhounds with more than 18 months of training were considered to be in veteran status.) The STU-100 was swept over multiple areas of the test subject, without contacting the subject, to obtain a volatiles profile not specific to any body part. In field trials, canines were offered the scent pads and they proceeded to follow the scent on 48-h-old trails. The veteran bloodhounds proved successful 96% of the time, while canines in the novice category had a 53% success rate.

Nations with human scent detection canine programs all utilize various natural fiber scent pad materials (e.g., cotton) to capture and release the volatiles profile. Although anecdotal and empirical data suggest that cotton materials collect and release organic chemicals, this study provides an organic chemical trap and release data confirmation from controlled laboratory experiments. Initial experiments focused on spiking the pads directly and utilizing a chromatographic separation before mass spectrometry for identification and confirmation. Additional experiments utilized a more direct analysis approach with no separation step before mass spectrometry. This latter approach generated multiple ion patterns or ion current profiles amenable to mathematical factor analysis. The pad used in the STU-100 sampling device was desorbed at ambient temperatures as well as elevated temperatures utilizing analytical instrumentation in controlled laboratory settings with known chemical mixtures or standards.

Experimental Methods

Matrix Spike Mix

Initial experiments were designed to determine the extent of desorption (release) of a known chemical mixture from the pads currently used with the STU-100. One microliter of a five-component, volatile organic analysis (VOA) Matrix Spike solution (EPA method 524.2, Volatiles in drinking water) prepared at 2.5 mg/mL in methanol was injected onto the inner cotton layers of a Johnson & Johnson$^{®}$ sterile dressing (currently used as the volatiles profile capture pad). The pad was immediately sealed inside a blanked (i.e., previously analyzed and found to contain nondetectable levels of the volatiles of interest), 3 L Tedlar^{TM} bag. The Tedlar^{TM} bag was then filled with 0.5 L of pure air ($\sim$ 25°C at 60% relative humidity) delivered by a Kin-Tek 491M Dynamic Gas Diluter (Kin-Tek Instruments, La Marque, TX). A negative control was made using the same procedure without addition of the VOA standard. A positive control was prepared by directly injecting 1 μL of the test mix into a blanked Tedlar^{TM} bag containing 0.5 L of pure air and no pad. The bags were allowed to equilibrate at room temperature for 1 h and 72 h, after which 10-mL headspace samples were removed from the bags by an Entech 7100 laboratory benchtop air concentrator (Entech Instruments, Simi Valley, CA). The air concentrator used three-stage preconcentration incorporating cryogen/Tenax TA solid sorbent, cryogen/glass beads, and a final cryofocusing stage before thermal injection into an Agilent 5973 Gas chromatography-mass spectrometry (GC/MS) fitted with a low thermal mass (LTM, RVM Scientific, Santa Barbara, CA) (17) DB-5 column (30 m $\times$ 0.25 mm OD $\times$ 0.25 μm) for analysis. The GC/MS method parameters are provided in Table 1. Quantitation of recovered analyte was based on a five-point calibration curve (0.25–25 ng/L) of the VOA mix generated before analysis.

Volatiles Gas Loading to the STU-100 Pad

To characterize adequately the adsorptive properties of the pads, a standardized NIST traceable gas mixture was used to simulate realistic sampling of low parts per billion by volume (p.p.b.$_v$) levels. A U.S. EPA compendium toxic organic compound (TO-14A) gas mixture (prepared at 1 part per million by volume [p.p.m.$_v$]) containing 39 target analytes was used as the test mixture (Restek Corp., Bellefonte, PA). According to Vass et al. (18) 14 of the 39 analytes in this mix have been identified as major components in decomposition of 1-year-old buried human remains. The implementation of this TO-14A gas mixture with this study will therefore also be beneficial to future scent determination research in conjunction with cadaver detection canines.

For the analysis (performed in triplicate), a volatiles capture pad was placed onto the STU-100, which was then positioned to face a $\frac{1}{8}$ in. diameter TO-14A gas stream situated 0.5 cm from the middle of the pad (Fig. 1). A diluted, humidified, 20 p.p.b.$_v$ TO-14A gas mixture (at 25°C) was then loaded onto the pad at a flow rate of 0.5 L/min for 2.0 min (20 ng nominal mass loading per analyte), during which time the STU-100 was operating at full power ($\sim$ 300 L/min). After gas loading, the scent pads were individually sealed inside blanked, 3 L Tedlar^{TM} bags and filled with 0.5 L of humidified pure air (25°C). The pads were allowed to equilibrate inside the bags for 3 h. Analysis consisted of concentrating the headspace (0.5 L) via the Entech laboratory air concentrator, followed by thermal desorption onto a GC-MS for TO-14A target analyte identification and quantitation. Quantitation of recovered analyte was based on a four-point calibration curve (0–20 ng) of the TO-14A gas mixture performed before analysis. This process was repeated with undiluted (1 p.p.m.$_v$) TO-14A gas mixtures as well.

TABLE 1—*Agilent GC/MS with LTM A68 method parameters.*

Agilent 6890 GC Oven	
Temperature	210°C
LTM-GC temperature program	
Initial temperature	30°C
Initial hold time	2.00 min
Temperature rate 1	20°C/min
Intermediate temperature 1	175°C
Hold time	0.00 min
Temperature rate 2	60°C/min
Final temperature	210°C
Final hold time	0.00 min
Transfer line/inlet	
Mode	Splitless
Temperature	250°C
Carrier gas	Helium
Column program	
Mode	Programmed pressure
Initial pressure	28.00 psi
Initial hold time	2.00 min
Pressure rate 1	2.02 psi/min
Intermediate pressure 1	38.11 psi
Hold time	0.00 min
Pressure rate 2	6.01 psi
Final pressure 2	46.10 psi
Final hold time	0.00 min
Average linear velocity	53 cm/sec
Agilent 5973 MSD settings	
Acquisition mode	Scan
Scan range	29–180 *m/z*, 0–1.5 min 34–280 *m/z*, 1.5–11 min
Scan rate	14.64 scans/sec, 0–1.5 min 10.13 scans/sec, 1.6–9.5 min
Threshold	150
Sampling (2^n)	$n = 1$
Solvent delay	0.00 min
MS quad/source temperature	150°/230°C
MS transfer line temperature	200°C

GC/MS, gas chromatography-mass spectrometry; LTM, low thermal mass.

Thermal Desorption with Atmospheric Chemical Ionization Mass Spectrometry (TD/APCI-MS)

A different analytical technique was utilized for analysis of the volatiles capture pads to provide additional supportive data for this evaluation. The alternate method used TD/APCI-MS. This technique offers the ability to rapidly obtain rapidly a mass spectrometric profile (or the volatiles profile in an ion current form) from analytes present on the pad that would ionize by this method. By using multivariate analysis on the acquired data, the reproducibility of the volatiles profiles can be compared and important discriminators can be determined.

A PE Sciex 365 triple quadrupole mass spectrometer (MDS Sciex, Concord, ON, Canada) equipped with a thermal desorption atmospheric pressure chemical ionization source (Mass Spec Analytica Inc., U.K.) was used for all of the experiments involving

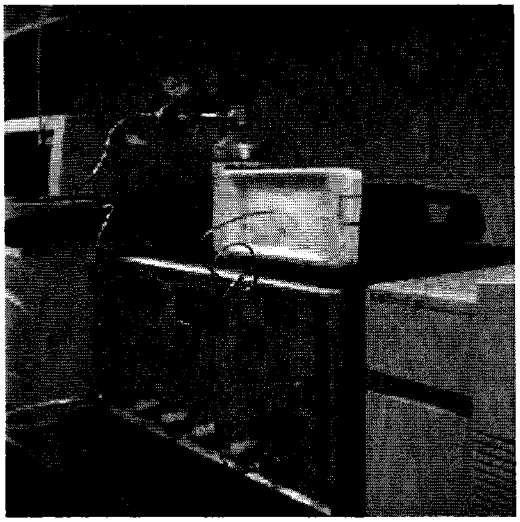

FIG. 1—*Scent Transfer Unit™ (STU-100) positioned to accept a TO-14A gas stream from the Kin-Tek 491M Dynamic Gas Diluter.*

direct analysis with no prior chemical separation of the components emanating from the pads. Figure 2 illustrates how the pad was inserted into the desorption block of the TD/APCI-MS for direct analysis. The pads were used with, or without, the STU-100 to collect volatiles from the headspace of selected chemicals. The pads were desorbed at ambient temperature or elevated temperatures to assist with the release process and mass spectra were collected from *m/z* 30 to 500. The resultant data were smoothed, centroided, and imported into a multivariate analysis software package (Pirouette Lite Explore v. 3.11) where principle component analysis (PCA) and hierarchical cluster analysis (HCA) were performed.

Experiments were conducted to demonstrate that chemicals were being adsorbed, absorbed into, or trapped onto the volatiles capture pads, and that the chemicals were released from the pads at room temperature. Approximately 1 mL of benzaldehyde was placed in a beaker and covered with a watch glass. An uncapped bottle of 2-nonenal was also placed in a beaker and covered with a watch glass. The headspace of each was sampled with the STU-100 for 5, 15, and 30 sec, with about 6–8 h between sampling. Only single measurements were acquired and the first 10 scans, unless otherwise noted, were averaged to yield the average ion current for selected compounds. In negative ion mode, a few milliliters of isovaleric acid was placed in a chamber and covered with a glass plate. The compound 2-ethylhexanoic acid was placed in a beaker and also covered with a watch glass. Sampling was performed for periods of 5, 15, and 30 sec. A binary mixture composed of approximately equal volumes of isovaleric acid and 2-ethylhexanoic acid was also prepared in a beaker and covered with a watch glass. Only semiquantitative analysis has been achieved via this technique thus far.

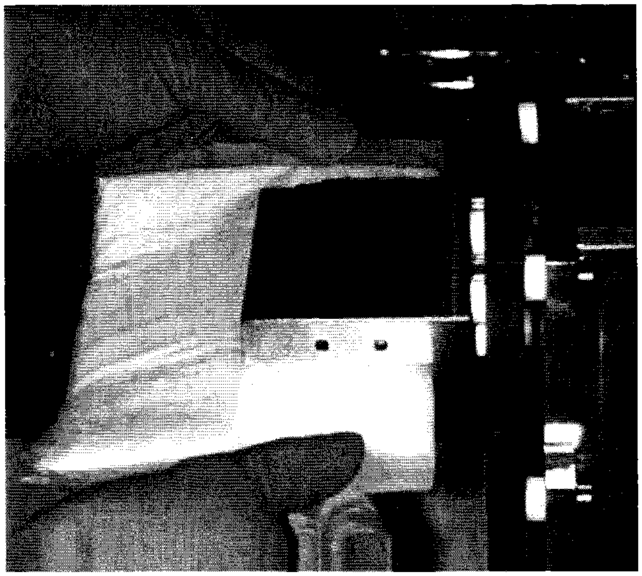

FIG. 2—*The volatiles capture pad was placed into the desorption block of the thermal desorption followed by atmospheric pressure chemical ionization mass spectrometry (TD/APCI-MS) for mass spectral analysis of the components that were successfully trapped by the pad.*

An example of the experimental data sets that can be generated via the TD/APCI-MS technique is shown in Fig. 3. In this particular data set, multiple ions are monitored as the analysis is executed. An example of two compounds that are reported to be human odor components, isovaleric acid and 2-nonenal, and their respective mass spectral patterns and selected ion profiles resulting from a headspace analysis, are illustrated. This experiment shows that when operating under these conditions, the molecular ions or the protonated molecules are primarily observed.

Results and Discussion

VOA Matrix Spike Mix

The results listed in Tables 2 and 3 indicate that volatile chemicals were released from the capture pad(s) and could be detected in a volume of as little as 10 mL of air at 1 and 72 h after loading. The positive controls resulted in a moderate recovery for four of the five analytes studied, with a high-recovery observed for benzene in the 1-h equilibration and for toluene in the 72-h equilibration. Although the positive control components at 72-h sampling exhibit, in general, higher recoveries than at the 1-h sampling, this was not observed in the actual pad samples. The recoveries in all three sampled scent pads were less at the 72-h mark than at the 1-h sampling (except for trichloroethylene in Sample #2), indicating that analyte may be readsorbing onto the pads over this time period, efficiently adhering to the inner surface of the Tedlar™ bag, or permeating through the bag. Nevertheless, this study indicates that measurable amounts of chemicals injected or spiked onto the scent pad in liquid form can be released over a 3-day period of time by the scent pad fibers into the headspace surrounding the pad.

Toxic Organics (TO-14A) Gas Mixture

The results along with the relative standard deviations (standard deviation/mean expressed as a percent [RSD]) are listed in Table 4 and show that a maximum of 15 out of the 39 target analytes were quantitated above the minimum detectable level (0.01 ng) when the scent pads were loaded at a concentration of 10 p.p.b.$_v$ (20 ng nominal mass). Most of these analytes detected were the heavier substituted aromatics, with the exception of two types of freons (dichlorotetrafluoroethane and trichlorofluoromethane), and 1,1-dichloropropane. Total recoveries for the remaining analytes were well below the load amount, suggesting three possibilities: (a) the pads were efficient in trapping the volatiles, but the flow rate (~ 300 L/min) of the STU-100 was too high to enable both efficient and reproducible adsorption of analyte to the pad, and/ or (b) the pads are only marginally efficient in trapping or

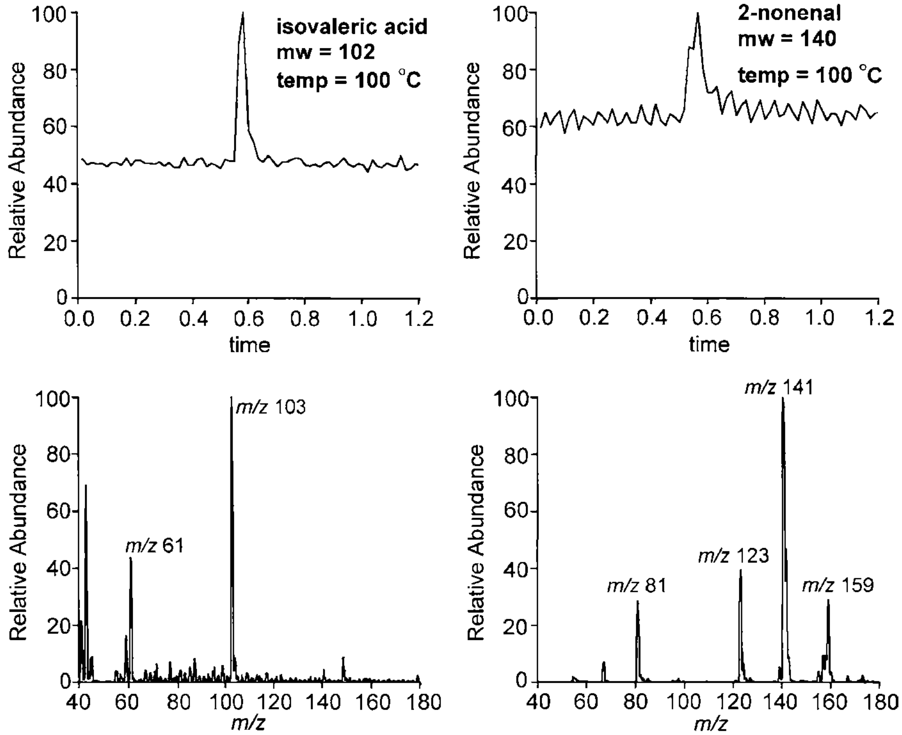

FIG. 3—*Example of the data generated by the thermal desorption followed by atmospheric pressure chemical ionization mass spectrometry (TD/APCI-MS) instrument when a volatiles capture pad is independently loaded with isovaleric acid and 2-nonenal. The total ion current (top two plots) and the individual spectra (bottom two plots) for the protonated molecule of isovaleric acid and 2-nonenal are illustrated at m/z 103 and 141, respectively.*

adsorbing the low-molecular-weight volatiles, and because of this inefficiency, analyte broke through the pads during sample loading, or (c) the pads required higher temperatures (>25–30°C) to release analyte in amounts sufficient for laboratory-based instrumentation to detect. To test the possibility of (a) above, sample loading was performed at the undiluted concentration (1 p.p.m.$_v$) of the TO-14 test mixture to load a nominal mass of 100 ng onto each scent pad, with one analysis incorporating the vacuum of the STU-100 and another analysis using only the undiluted gas stream (0.1 L/min) onto the pad. The results along with the calculated

TABLE 2—*VOA matrix spike results after 1-h equilibration.**

Target Compound	Theoretical Yield (ng)	Analyte Recovery							
		Control		Sample #1		Sample #2		Sample #3	
		ng	%	ng	%	ng	%	ng	%
1,1-Dichloroethene	12.76	2.65	21	0.12	0.9	0.11	0.9	0.2	1.6
Benzene	15.6	13.91	89	1.32	8.5	0.67	4.3	5.64	36
Trichloroethylene	9.4	1.12	12	0.73	7.8	0.62	0.7	2.38	25
Toluene	13.28	3.35	25	2.45	18	2.42	18	6.7	50
Chlorobenzene	10.92	3.14	29	1.57	14	1.99	18	3.09	28

*All blanks yielded nondetects.
VOA, volatile organic analysis.

TABLE 3—*VOA matrix spike results after 72-h equilibration.**

Target Compound	Theoretical Yield (ng)	Analyte Recovery							
		Control		Sample #1		Sample #2		Sample #3	
		ng	%	ng	%	ng	%	ng	%
1,1-Dichloroethene	12.76	3.85	30	ND	—	ND	—	0.17	1.3
Benzene	15.6	5.51	35	0.15	0.9	0.42	2.7	1.7	11
Trichloroethylene	9.4	5.29	56	0.14	1.6	1.05	11	1.1	12
Toluene	13.28	14.93	112	0.39	2.9	1.63	12	2.85	21
Chlorobenzene	10.92	3.35	31	0.52	4.8	0.9	8.2	1.06	9.7

*All blanks yielded nondetects (ND).
VOA, volatile organic analysis.

TABLE 4—*Volatiles profile capture pad results for 10 p.p.b.$_v$ gas load during STU-100 operation.*

Target Compound^{++}	Recovery (ng)*				
	Control^{+}	Sample Pad #1	Sample Pad #2	Sample Pad #3	RSD (%)
Dichlorotetrafluoroethane†	15.84	0.13	0.2	0.22	26
Trichlorofluoromethane†	18.74	0.23	0.04	ND	100
1,2-Dichloropropane	22.41	0.6	ND	ND	—
Toluene†	29.96	ND	0.2	0.07	68
Ethylbenzene†	15.29	ND	0.29	0.3	2
m-Xylene†	15.64	ND	0.26	0.27	3
p-Xylene†	15.64	ND	0.26	0.27	3
Styrene†	15.84	ND	0.15	ND	—
o-Xylene†	15.74	ND	0.15	0.25	35
1,3,5-Trimethylbenzene	19.91	0.04	0.04	0.02	35
1,2,4-Trimethylbenzene	19.3	ND	0.04	0.02	47
1,4-Dichlorobenzene	9.98	0.06	0.05	ND	13
1,2-Dichlorobenzene	9.86	0.05	0.08	0.11	38
1,2,4-Trichlorobenzene	4.42	0.14	0.15	0.14	4
Hexachlorobutadiene	11.36	0.04	0.04	0.02	35

*Most analytes from the TO-14A gas mix were below minimum detection level (<0.01 ng).

†Denotes compounds identified as major components in 1-year-old buried human remains (16).

‡Nominal 20 ng analyte loading.

§All blanks yielded nondetects (ND).

STU-100, Scent Transfer Unit™.

RSDs are listed in Tables 5 and 6, respectively. The analytes and recoveries at the higher concentration of 1 p.p.m.$_v$ using the STU-100 at a high flow rate (Table 5) coincided with those from the 10 p.p.b.$_v$ study, whereas higher recoveries of a larger number of analytes were obtained by loading the scent pads without the use of a high flow (Table 6). These results suggest that although the recoveries remain comparatively low against the theoretical load, the high flow of the STU-100 may reduce the adsorptive and/or absorptive efficiency of the scent pads, resulting in lower detectable amounts. Nevertheless, the pads demonstrated the ability to *trap* a limited number of gaseous analytes from a 0.1 L/min gas stream and *release* quantities at ambient temperature into the surrounding air at levels detectable by analytical laboratory instrumentation.

TABLE 5—*Volatiles profile capture pad results for 1.0 p.p.m.$_v$ gas load during STU-100 high-flow operation.*

Target Compound^{++}	Recovery (ng)*+			
	Sample Pad #1	Sample Pad #2	Sample Pad #3	RSD(%)
Dichlorotetrafluoroethane†	0.23	0.23	0.24	2
Trichlorofluoromethane†	0.21	0.21	0.21	0
1,2-Dichloropropane	ND	1.01	1.4	23
Toluene†	0.18	0.2	ND	7
Ethylbenzene†	0.29	0.28	0.28	2
m-Xylene†	0.25	0.25	0.24	2
p-Xylene†	0.25	0.25	0.24	2
Styrene†	0.13	0.13	0.15	8
o-Xylene†	0.15	0.15	0.15	0
1,3,5-Trimethylbenzene	0.06	0.02	0.02	69
1,2,4-Trimethylbenzene	0.1	ND	ND	—
1,4-Dichlorobenzene	0.23	0	0.05	94
1,2-Dichlorobenzene	0.09	0.05	ND	40
1,2,4-Trichlorobenzene	0.21	0.15	0.13	25
Hexachlorobutadiene	0.09	0.02	0.02	93

*Most analytes from the TO-14A gas mix were below minimum detection level (<0.01 ng).

†Denotes compounds identified as major components in 1-year-old buried human remains (16).

‡Nominal 100 ng analyte loading.

§All blanks yielded nondetects (ND).

STU-100, Scent Transfer Unit™.

TABLE 6—*Volatiles profile capture pad results for 1.0 p.p.m.$_v$ gas load under no flow STU-100 conditions.*

Target Compound^{++}	Recovery (ng)$^{+}$			
	Sample Pad #1	Sample Pad #2	Sample Pad #3	RSD (%)
Dichlorodifluoromethane†	ND	0.1	ND	—
Chloromethane	ND	0.55	0.38	26
Dichlorotetrafluoroethane†	0.12	0.23	0.12	41
Trichlorofluoromethane†	0.29	0.22	0.28	14
1,1-Dichloroethene	ND	0.13	ND	—
Trichlorotrifluoroethane	0.24	0.21	0.22	7
(Z)-1,2-Dichloroethene	0.14	0.15	0.13	7
Chloroform†	0.05	0.06	0.04	20
1,2-Dichlrorethane	0.08	0.11	0.05	38
Carbon tetrachloride†	0.03	0.05	0.09	54
Benzene†	0.12	0.09	0.09	17
Trichloroethylene†	0.07	0.13	0.06	44
(Z)-1,3-Dichloropropene	ND	0.15	0.14	5
1,1,2-Trichloroethane	0.07	0.07	0.05	18
Toluene†	0.23	0.19	0.22	10
Chlorobenzene	0.34	0.4	0.32	12
Ethylbenzene†	0.33	0.37	0.31	9
m-Xylene†	0.3	0.35	0.28	12
p-Xylene†	0.3	0.35	0.28	12
Styrene†	0.24	0.31	0.19	24
1,1,2,2-Tetrachloroethane	0.28	0.33	0.19	27
o-Xylene†	0.15	0.16	0.15	4
1,3,5-Trimethylbenzene	0.23	0.35	0.12	49
1,2,4-Trimethylbenzene	0.28	0.82	0.17	82
1,3-Dichlorobenzene	0.31	1.12	0.15	99
1,4-Dichlorobenzene	0.62	1.73	0.35	81
1,2-Dichlorobenzene	0.44	1.15	0.28	74
1,2,4-Trichlorobenzene	0.78	2.1	0.6	71
Hexachlorobutadiene	0.69	2.06	0.53	77

*Nominal 100 ng analyte loading.

†Denotes compounds identified as major components in 1-year-old buried human remains (16).

‡All blanks yielded nondetects (ND).

STU-100, Scent Transfer Unit™.

Thermal Desorption Atmospheric Chemical Ionization Mass Spectrometry Analysis

Two chemicals that have been identified as potentially contributing to human odor were selected to evaluate alternatively the volatiles capture pads' ability to trap and release the material being sampled: benzaldehyde and 2-nonenal. The headspace over benzaldehyde, a possible odor component from hands (9), and 2-nonenal, a scent component associated with people less than 40 years of age (19), was sampled with the STU-100. The positive ion mode single measurement results are shown in the bar graphs in Figs. 4, 5*a* and *b*. The average ion signal intensity for the protonated molecule of benzaldehyde at *m/z* 107 is plotted in Fig. 4. Figure 5*a* shows the average ion signal from the protonated molecule of 2-nonenal (*m/z* 141) as a function of sampling time; Fig. 5*b* shows the average ion intensity summing the ion current generated from the protonated molecule as well as two fragment ions, *m/z* 123 and 81 from 2-nonenal. These results indicate the ability of the pad to capture chemical compounds associated with

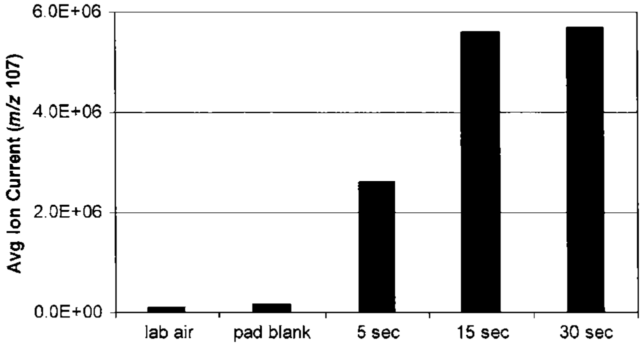

FIG. 4—*Averaged ion signal (eight scans) for the protonated molecule of benzaldehyde as a function of sampling time (and controls) with the Scent Transfer Unit^{TM} (STU-100) at room temperature.*

human odor using the STU-100 device and that these compounds can desorb from the pads at room temperature.

In negative ion mode, isovaleric acid, a foot malodor component (20), and a component found in underarm odor, 2-ethylhexanoic acid (21), were sampled using the STU-100. A binary mixture composed of *c.* equal volumes of isovaleric acid and 2-ethylhexanoic acid was also sampled for various time periods. The experimental results are shown in Figs. 6 and 7. The negative ion data, as with the positive ion data, show that chemicals are being adsorbed onto the pad and released at room temperature.

In Figs. 4–7, it is important to note that laboratory air and "blank" pads generate a measurable, albeit low, ion current via this technique. Figures 4–6 are plotted with absolute ion current on the ordinate to illustrate that "blank" pad levels for the analytes of interest are relatively low, yet present. Ideally, any pad used for evidence collection would have a nonmeasurable background level; however, in reality, this is difficult to achieve. Nevertheless, SFE has been shown to generate analytically "clean" pads to reduce and/or eliminate background levels present on virgin pads (22). The protocol for using the STU with canines mandates the use of a control presented to the canine to establish a local/scene background before sampling evidence with the volatiles capture pad and to establish a negative response to trail.

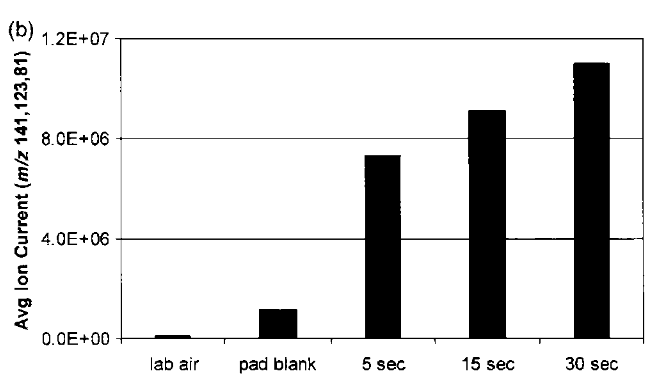

FIG. 5—*(a) Averaged ion signal (10 scans) for the protonated molecule of 2-nonenal as a function of sampling time (and controls) with the Scent Transfer Unit^{TM} (STU-100) at room temperature. (b) Averaged ion signal (10 scans) summing the protonated molecule of 2-nonenal and fragment ions at m/z 123 and 81 plotted versus sampling time (and controls).*

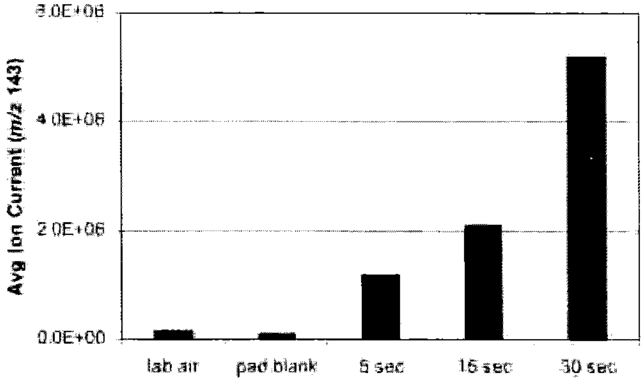

FIG. 6—*Averaged ion signal (10 scans) for the protonated molecule of 2-ethylhexanoic-acid as a function of sampling time (and controls) with the Scent Transfer Unit^{TM} (STU-100) at room temperature.*

Statistical Analysis

To further supplement the experimental data, statistical analysis was performed to provide mathematical visualization for the delineation of volatiles profile patterns, especially when using the rapid technique of TD/APCI-MS. Mass spectral data acquisition via TD/APCI-MS lends itself to multivariate analysis due to the multiple ion monitoring or detection that is used. In each analysis, a rapid desorption yields ion current as *m/z* measurements from all of the analytes present that can be ionized by APCI; however, using multivariate analysis or PCA, patterns within the data that arise due to the variances can be readily observed by using three-dimensional graphing.

PCA of the volatiles profile data demonstrates reproducibility achievable with the STU-100, as well as determines important discriminators in identifying individuals. PCA of unused pads showed variability between boxes; however, the data cluster together, as illustrated in Fig. 8. Replicate samples of pads used to sample the headspace of isovaleric acid (commonly observed emanating from "sweaty socks"), and those samples acquired directly over spiked isovaleric scent pads, showed similar clustering as well. Also clearly differentiated from the isovaleric volatile profiles were the pads spiked with heptanoic acid. This single plot shows clear instances of the clustering observed between isovaleric acid, heptanoic acid, background air, and virgin pads,

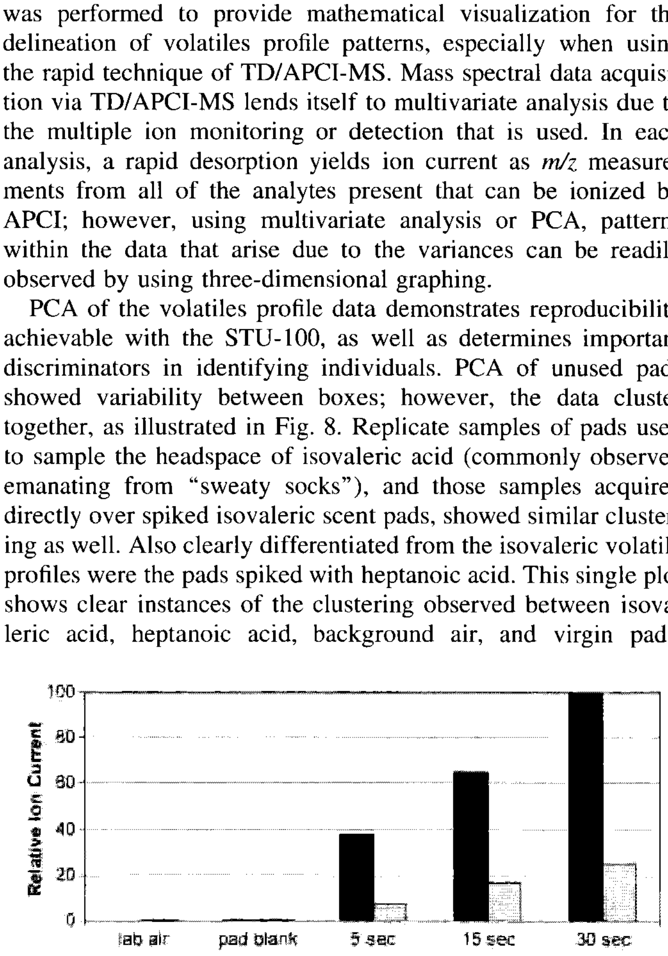

FIG. 7—*Averaged ion signal (10 scans) for the two components in a binary mixture of isovaleric acid (hatched) and 2-ethylhexanoic-acid (textured) sampled using the Scent Transfer Unit^{TM} (STU-100) as a function of sampling time (and controls) at room temperature.*

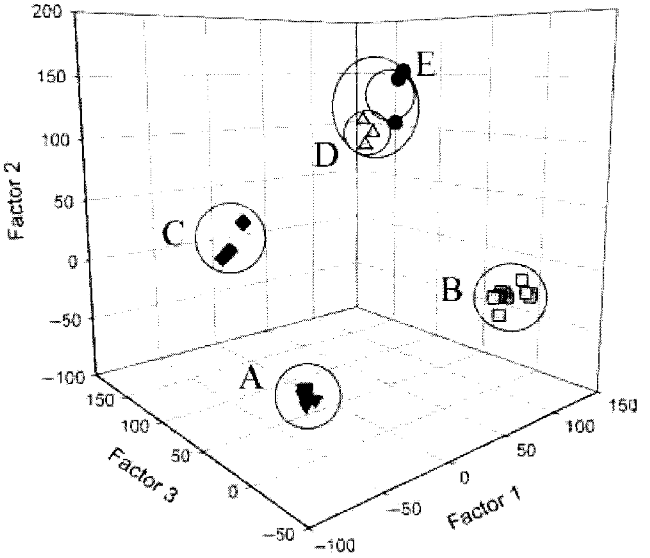

FIG. 8—*Principle component analysis (PCA) scores plot showing three factors that assist in the delineation between (A) background air, (B) virgin volatiles capture pads, (C) a pad spiked with heptanoic acid, (D) a pad loaded by sampling the headspace over a few milliliters of neat isovaleric acid, and (E) a pad spiked with isovaleric acid.*

collected using the TD/APCI-MS technique in negative ion mode. The source of the variability and thus an important discriminator for the isovaleric acid pads as identified in the PCA loading plot (not shown) was predominantly *m/z* 101 and 203, corresponding to the $(M{-}H)^-$ and $(2M{-}H)^-$ ions of isovaleric acid, respectively.

To demonstrate that the volatiles capture pads can adsorb complex mixture components and to show that PCA can help delineate these upon release or desorption, a variety of various colognes were sampled via the STU-100. The mass spectral ion current patterns resulting from the TD/APCI-MS analysis (50°C) of the pads after sampling the vapors emanating from four different colognes are shown in Fig. 9. Although the mass spectral patterns between virgin pads and the Aramis cologne are similar visually, the PCA analysis can delineate the two as shown in Fig. 10. The other complex colognes can also be distinguished in factor space. These data lend support to the fact that perhaps canines could differentiate these volatiles (scent/odor) profiles.

Conclusions

In the analytical literature, there are a host of sampling methods for measuring a wide variety of chemicals in a manner that is scientifically sound, reliable, and defensible. Before general acceptance of any sampling method, controlled testing is performed to ascertain the strengths and weaknesses of the method or the equipment used to make the measurements. In this instance, the STU-100 sampling device was evaluated to determine whether the cotton pads used in the device actually trap volatile or semivolatile chemicals and release them at room temperature. What may seem obvious to the casual observer given empirical

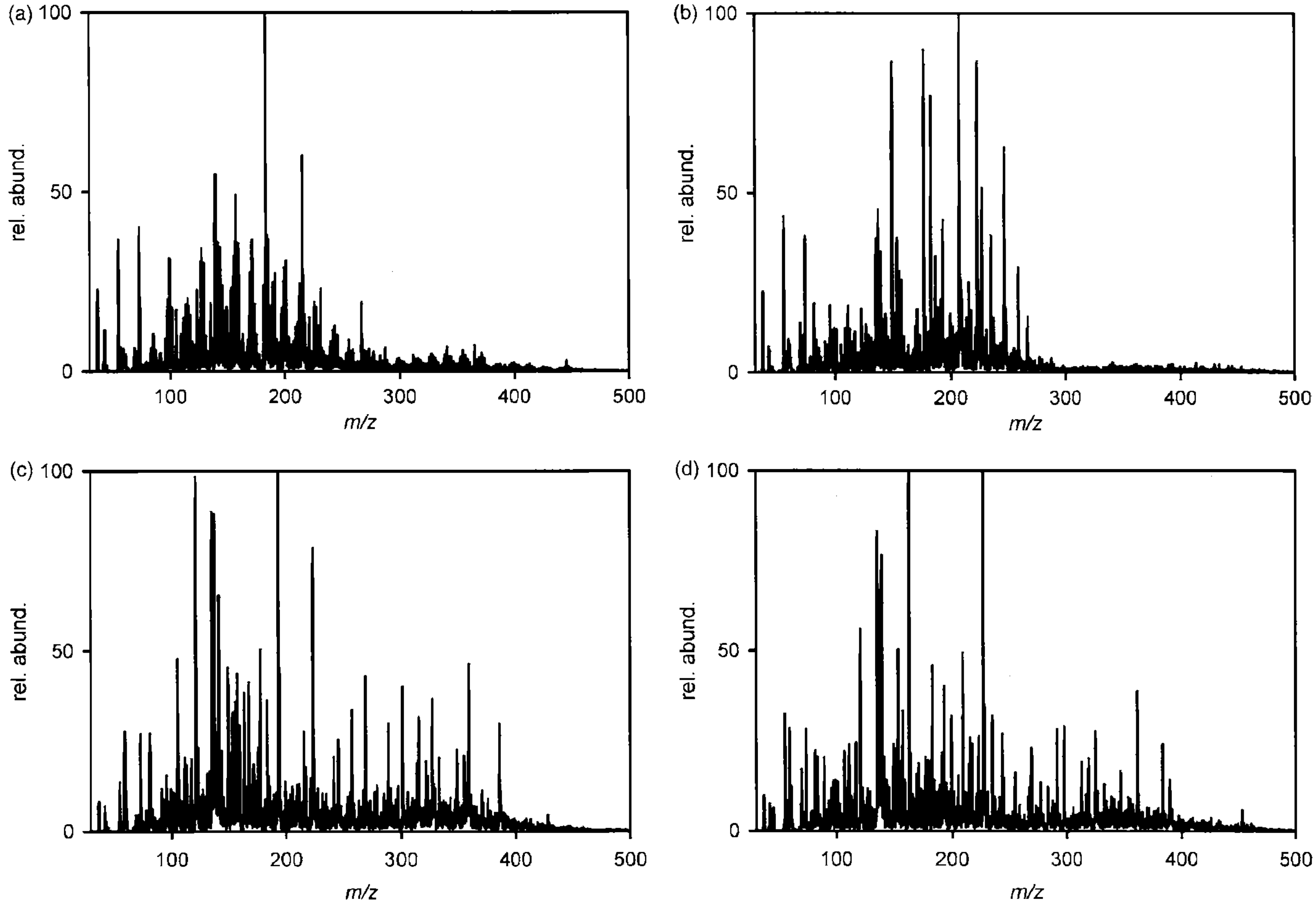

FIG. 9—*Normalized ion current (ordinate) thermal desorption followed by atmospheric pressure chemical ionization mass spectrometry (TD/APCI-MS) profiles resulting from desorption of the Scent Transfer Unit™ (STU-100) volatiles profile capture pads subsequent to sampling a variety of complex mixtures. The mass spectral ion current profiles are (A) virgin pad, (B) Davidoff cologne, (C) Tea Rose cologne, (D) Tommy cologne, and (E) Aramis cologne.*

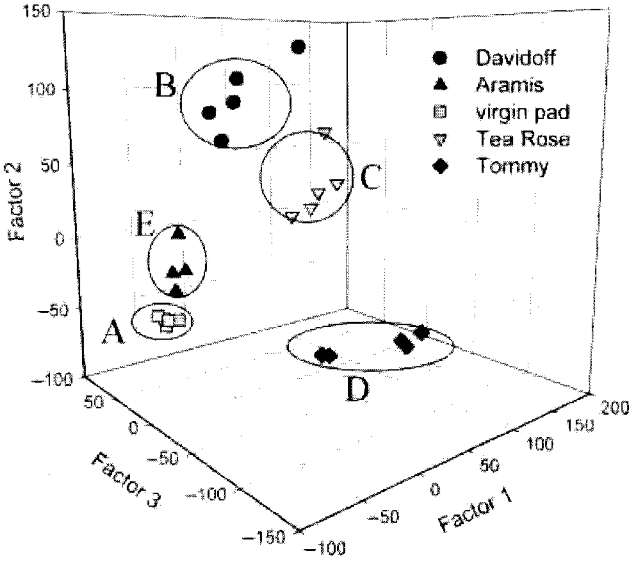

FIG. 10—*Principle component analysis (PCA) three-dimensional scores plot resulting from statistical analysis of the data generated from the thermal desorption followed by atmospheric pressure chemical ionization mass spectrometry (TD/APCI-MS) experiment utilizing four different colognes sampled by the Scent Transfer Unit™ (STU-100). Each of the different colognes cluster as: (A) virgin pad, (B) Davidoff, (C) Tea rose, (D) Tommy, and (E) Aramis.*

knowledge of cotton materials, a controlled scientific investigation into verifying these assumptions for legal precedence was still required for this device. The fact that the cotton pads used in the STU-100 sampling device can trap and release chemicals at ambient temperature has been experimentally verified. Although the trapping and release efficiencies were found to be less than optimal, these experiments were required to characterize the device and to establish a baseline for improving it.

In each of the two analytical methodologies, GC/MS and TD/APCI MS, the STU-100 performance evaluation experimental test plan consisted of a strategy using simple chemical spikes followed by more complex mixtures. With the GC/MS experiments, testing moved from a simple liquid spike mixture to a more complicated, 39-component TO-14A gaseous mixture. With the TD/APCI MS experiments, analytical measurements were performed from simple and binary mixtures as well as complex organic mixtures illustrated via the cologne analyses. Each analytical methodology used demonstrated effective loading and subsequent release of analytes with high volatility, in both liquid and gaseous form. In addition, the multivariate analysis yielded insights into the volatile and semivolatile organic compound patterns and clustering that can be differentiated in factor space. The mass spectral patterns that resulted after spiking the volatiles capture pads with various perfumes can be distinguished by the TD/APCI-MS approach both visually as well as chemometrically; however, by displaying the data in factor space, clustering of the patterns is evident and lends further support to the fact that the pads utilized in the STU-100 can trap (adsorb) and release (desorb) chemical mixtures.

This research has spawned many new ideas for adjusting, modifying, improving, and exploring further the advantages or disadvantages of the STU-100. Future directions will involve the use of test mixtures that are representative of the human odor components that have been reported experimentally (9,10) to provide a better understanding of the nature of canine olfactory detection and to exploit variation of airflow volumes for more efficient volatiles collection (adsorption). Research to improve the pad capture capabilities via surface modifications, new polymeric trapping materials, and flow modulation will be necessary to increase the levels of analyte collected for the canines as well as for laboratory analytical instrumentation. Further statistical analysis will be required to illustrate whether clusters formed in the multidimensional factor analysis plots actually can assist scientists in determining uniqueness among individuals or assist in the differentiation of individuals or groups and whether these clusters correlate with trailing canine volatiles profile (scent/odor) experiments. Finally, continued volatiles sampling of humans and potential items of evidence will be necessary to further evaluate or validate the analytical protocols and improve the overall understanding of the complexities of human volatiles profile evidence sampling and recovery.

Acknowledgments

The authors gratefully acknowledge John Luke of Mass Spec Analytica for loaning the thermal desorption APCI unit through a Cooperative Research and Development Agreement (CRADA #ORNL02-0662). S. Ramsey acknowledges support in part through the Federal Bureau of Investigation's Visiting Scientist Program, an educational opportunity administered by Oak Ridge Institute for Science and Education (ORISE).

References

1. Daubert V. Merrell Dow Pharmaceuticals. 509 U.S. 579; 113 S. Ct. 2786, 1993.
2. Frye V. United States, 54 App. D.C. 46; 293 F. 1013, 1923.
3. Stockham R, Slavin D, Kift W. Specialized use of human scent in criminal investigations. Forensic Science Communications 2004;6(3): http://www.fbi.gov/hq/lab/fsc/backissu/july2004/research/2004_3_research03.htm.
4. Taslitz A. Does the cold nose know? The unscientific myth of the dog scent lineup. Hastings Law J 1990;42:15–134.
5. Settle R, Broom D, Sommerville B, Darling F. The use of trained dogs to discriminate human scent. Anim Behav 1993;46:189–90.
6. Wojcikiewicz J. Dog scent lineup as scientific evidence. Int Acad Forensic Sci 1999.
7. Peterson J, Markham P. Crime laboratory proficiency test results, 1978–1991, II: resolving questions of common origin. J Forensic Sci 1995;40(6):1009–29.
8. Furton K, Lorenzo N, Wan T, Harper R, Chow M, Rose S. Laboratory and field experiments used to identify Canis lupus var. familiaris active odor signature chemicals from drugs, explosives, and humans. Anal Bioanal Chem 2003;376(8):1212–24.
9. Bernier U, Kline D, Barnard D, Schreck C, Yost R. Analysis of human skin emanations by gas chromatography/mass spectrometry. 2. Identification of volatile compounds that are candidate attractants for the yellow fever mosquito (*Aedes aegypti*). Anal Chem 2000;72(4):747–56.
10. Benner B, Goodpastor J, DeGrasse J, Tully L, Levin B. Characterization of surface organic components of human hair by online supercritical fluid extraction-gas chromatography/mass spectrometry: a feasibility study and comparison with human identification using mitochondrial DNA sequences. J Forensic Sci 2003;8(3):554–63.
11. Nicolaides N. Skin lipids: their biochemical uniqueness. Science 1974;186:19–26.
12. Curran A, Rabin S, Furton K. Analysis of the uniqueness and persistence of human scent. Forensic Sci Commun 2005;7(2):1–21.
13. Morrison Ed. Canine olfactory systems structure and function. Proceedings of the Olfactory Based Systems for Security Applications (OBSSA). London, UK: DSTL, 2004.
14. Tolhurst W, Harris L. US Patent #D397051, August 18, 1998.
15. Shane S. FBI's use of bloodhounds in anthrax probe disputed. Baltimore Sun, 2002, http://www.ph.ucla.edu/epi/bioter/fbisuseofbloodhounds.html
16. Harvey L, Harvey J. Reliability of bloodhounds in criminal investigations. J Forensic Sci 2003;48(4):1–6.
17. Sloan K, Mustacich R, Eckenrode B. Development and evaluation of a low thermal mass gas chromatograph for rapid forensic GC-MS analysis. Field Anal Chem Technol 2001;5(6):288–301.

18. Vass A, Smith R, Thompson C, Burnett M, Wolf D, Synstelien J, et al. Decompositional odor analysis database. J Forensic Sci 2004;49(4):1–10.
19. Haze S, Gozu Y, Nakamura S, Kohno Y, Sawano K, Ohta H, et al. 2-nonenal newly found in human body odor tends to increase with aging. J Invest Dermatol 2001;116(4):520–4.
20. Kanda F, Yugi E, Fukuda M, Nakajima K, Ohta T, Nakata O. Elucidation of chemical compounds responsible for foot malodour. Br J Dermatol 1990;122:771–6.
21. Zeng X, Leyden J, Brand J, Spielman A, McGinley K, Preti G. An investigation of human apocrine gland secretion for axillary odor precursors. J Chem Ecol 1992;18(7):1039–55.
22. Curran A, Rabin S, Prada P, Furton K. Comparison of the volatile organic compounds present in human odor using SPME-GC/MS. J Chem Ecol 2005;31:1607–18.

Additional information and reprint requests:
Brian A. Eckenrode, Ph.D.
FBI Academy
Building 12
Quantico, VA 22135
E-mail: baeckenrode@fbiacademy.edu

APPENDIX B

Available online at www.sciencedirect.com

SCIENCE DIRECT

Forensic Science International 147 (2005) 43–47

Forensic Science International

www.elsevier.com/locate/forsciint

The effect of the ageing of crime scene objects on the results of scent identification line-ups using trained dogs

G.A.A. Schoon[a,b,*]

[a] *Animal Behaviour Group, Leiden University, P.O. Box 9516, 2300RA Leiden, The Netherlands*

[b] *Canine Department of the Netherlands National Police Agency, P.O. Box 530, 8070AM Nunspeet, The Netherlands*

Received 28 October 2003; received in revised form 30 March 2004; accepted 2 April 2004

Abstract

In a scent identification line-up, a trained dog matches the scent trace left by a perpetrator at the crime scene to the odour of a suspect in a line-up of different odours. The procedures are strictly defined and the results are routinely used by the police and as evidence in court in a number of European countries. This paper describes the effect of ageing of the odour trace collected at the crime scene on the performance of the dogs in recognising the perpetrator in a line-up. The results show that whilst the dogs perform faultlessly in matching odours collected on the same day, the results drop to a lower level and become more variable in the period studied (2 weeks to 6 months). However, the results do not show a systematic decrease in performance. A possible explanation is the development of a steady state in the glass jars containing the perpetrator odour trace after initial differential evaporation of components of the residue or break down of unsaturated components into saturated ones. Prevention of this initial change may prevent the drop in performance observed in this study, thus increasing the reliability of these scent identifications.

Keywords: Scent; Identification; Canine; Ageing

1. Introduction

A scent identification line-up is a procedure where a trained dog works in a match to sample set-up. Usually, the dog is given scent collected in some way at a crime scene as a sample, and is asked to compare this odour with a number of odours in a row, one of which is the odour of the suspect. Scent identification line-ups are performed in a number of European countries, and accepted as part of the evidence presented in court [1].

Research conducted in this field is slow work and fraught with difficulties. It takes a long time to train a dog, and therefore the numbers of dogs available at any time is low. The police who train the dogs prefer to use them for operational cases instead of keeping them available for research. The performance of the dogs can be variable. Sometimes a dog is retired early when it does not perform well. Collecting odours for research purposes is time consuming and it is sometimes difficult to find sufficiently large groups of people willing to co-operate in giving their odours. Learning effects can influence results if the same group of people is used in successive experiments. This has led to a situation where the results are accepted in court, but there are still questions about the scientific base and the reliability of the results [2].

Continuing research efforts are being made especially in Poland and in Holland. Much research focuses on the effect of variables such as contamination or masking and the possible use of other odour cues than the unique individual odour of a person. However, one basic question of these line-ups has not yet been addressed. In the majority of cases, the sample odour collected at the scene of the crime is collected some time before the possibly matching odour of the suspect. Given its volatile nature, odour traces change with time, as is known from perfumes. The chemical composition of residue on objects after handling has also been shown to change with time [3]. The effect of these changes caused by time on the performance of the dogs has not been studied yet.

---

[*] Tel.: +31 341 278900; fax: +31 341 262682.
E-mail address: afd.speurhonden@worldmail.nl (G.A.A. Schoon).

0379-0738/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2004.04.080

This paper describes two experiments conducted in The Netherlands. The first experiment followed a set-up designed to obtain a lot of information in a minimal time, the second experiment was conducted following the usual scent identification line-up protocol used by the police in The Netherlands.

2. Material and methods

2.1. *Experiment 1*

2.1.1. *Subject*

Dog A: German Shepherd cross breed, male, age 8 years.

2.1.2. *Material*

Odours were collected on three different kinds of material: stainless steel metal tube, PVC electric tubing, or cotton cloth.

Eight different males contributed target odours on the three kinds of material on eight occasions during half a year. They did this by keeping an object with them for 10 min in a pocket, and rubbing it well between their hands for a short while. In this way, 187 target odours were collected (on five occasions, the target person forgot). For each target odour, four similar foil odours were collected from different people at the same time and on the same kind of material as the target odour.

2.1.3. *Experimental set-up*

Five jars containing odours were placed in a circle. During a test, the dog had to search for the odour of a specific target person during nine trials. The first of the nine trials was a "priming" trial, where the dog was given the target odour amidst unscented pieces of similar material. In the subsequent eight trials, the target odour was presented among four similar foil odours. The position of the target odour was random. The age of the odours was increased in steps: 0, 2, 4, 8, 12, 16, 20 and 24 weeks. The dog handler did not know the position of the target odour, and signalled when the dog made its response (lying down, or standing over the jar of its choice). A correct choice was rewarded with a retrieving game (the handler would pretend the toy came out of the chosen jar, the usual training reward). After a no-recognition (the dog had smelled all of the jars at least once but did not react to any of the odours) or an incorrect choice (a full response as interpreted by the handler for a non-target odour), the first priming trial was repeated to refresh the memory of the dog and to standardise that each trial was preceded by a "successful" trial that had led to a reward.

2.2. *Experiment 2*

2.2.1. *Subjects*

Dogs B–K: Malinois and German Shepherd cross breeds, three females, seven males, varying in age from 2–7-year-old. Two of these dogs were operational German dogs, the others were operational Dutch dogs.

2.2.2. *Material*

To simulate odour traces left at a crime scene, 10 different target people were asked to scent 8 pieces of cotton cloth at time 0 by holding them in their pocket for some minutes, and rubbing them well between their hands. The cloths were packaged separately in glass jars with a twist off top. The odours for the line-up were collected in the official way at time 0, and 2, 4, 8, 12, 16, 20 and 24 weeks later. For each line-up, the target person and six others (foils) were asked to scent stainless steel tubes by holding them in their hands for approximately 2 min. One foil was also asked to scent a piece of cloth for the control trial in the protocol. The tubes (and the control piece of cloth) were packaged separately in glass jars with twist-off tops. This material was used within a week for the line-ups. Each of the dogs worked with the odour of a single target person.

2.2.3. *Experimental set-up*

The tubes were clamped in holder on two platforms, each platform containing the odours of the target person and the six foils, but in a different sequence as determined by dice. The protocol followed five steps:

1. Match the odour of the control cloth to the matching control foil odour in row 1;
2. Match the odour of the control cloth to the matching control foil odour in row 2;
3. Evaluate the interest for the dog for the target odours in the line-up;
4. Match the odour of the simulated crime scene cloth to the target person in row 1;
5. Match the odour of the simulated crime scene cloth to the target person in row 2.

For steps 1, 2, 4 and 5 the handler would stand in front of the row with his dog, and give the dog the sample scent. After a command "search", the dog would search for a matching odour on the platform. The handler did not know the position of the matching odour, and observed his dog. When his dog indicated its choice, the handler would raise his hand, and would be given a signal (red/green light) to inform him if the correct choice had been made. If the dog made a correct choice, the tube would be released from its holder, and the dog was allowed to run around with it for a while and bring it to its handler as a reward. After an incorrect choice, or if the dog did not make a choice at all, the handler would recall the dog and it would not be rewarded.

No choice or choosing a foil in steps 1 and 2 led to a disqualification. Special interest for the target odour (step 3) also led to a disqualification. Matching the simulated crime scene cloth to the target odour in both steps 4 and 5 was called correct recognition. Not responding to any odour in step 4 or 5 was a no-recognition, responding to a foil in these steps was an incorrect choice. If a dog was disqualified in a

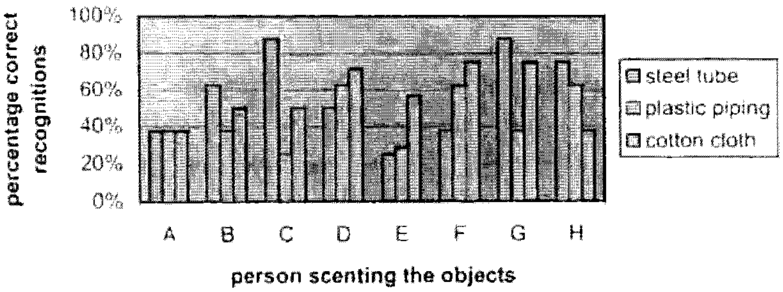

Fig. 1. Results of experiment 1: percentage correct recognitions of different material types of all ages scented by different people by dog A.

line-up, it was usually given a second chance later in that week, using a duplicate set of tubes.

3. Results

3.1. Experiment 1

The results were analysed in a number of different ways. First, the difference between the eight people and the material type were analysed. Here, all the different ages were taken together. These results are given in Fig. 1. Percentages correct are used since occasionally data was missing due to a person forgetting to scent the objects on time. On average the results varied from 37 to 67% correct choices between individuals, but no single individual was recognised systematically better (or worse) with all material types. In Table 1, the results are averaged per material type. None of the material types is recognised significantly better (or worse) than the others.

Table 1
Average percentage recognitions and standard deviations of different material types cumulated over persons and age by dog A

	Average (%)	S.D.
Steel tube	57.8	24.3
Plastic piping	44.2	15.2
Cotton cloth	56.7	15.5

Finally, in Fig. 2 the results are cumulated per material type in time. None of the material types is recognised systematically better or worse than the others. After the initial 100% recognition at time 0, recognition drops, and varies between 25 and 61% (average 45.8%, standard deviation 19.0).

3.2. Experiment 2

In total, 90 line-ups were conducted. In 14 line-ups, the dog was disqualified in either steps 1–3. In 10 of these cases, the line-up was repeated later on in the week, in the other 4 this was not possible due to lack of material. In one case, the second line-up also led to a disqualification. After a successful qualification, two dogs only chose correctly based on the time 0 target odour. The other dogs varied between five to seven correct choices out of eight. The combined results are given in Table 2. After the initial 100% recognition at time 0, recognition drops and varies between 33 and 75% (average 58.8%, standard deviation 13.6). If the results of the first two non-performers are disregarded, the average recognition rate after time 0 climbs to 73.7% (standard deviation 17.9).

3.3. Combined results experiment 1 and 2

The results of the cloth-trials of experiment 1 are combined with the results of experiment 2 (excluding the non-performers) in Fig. 3. After the initial 100% recognitions at time 0, recognition drops but the operational dogs perform better on average than dog A of the experiment 1.

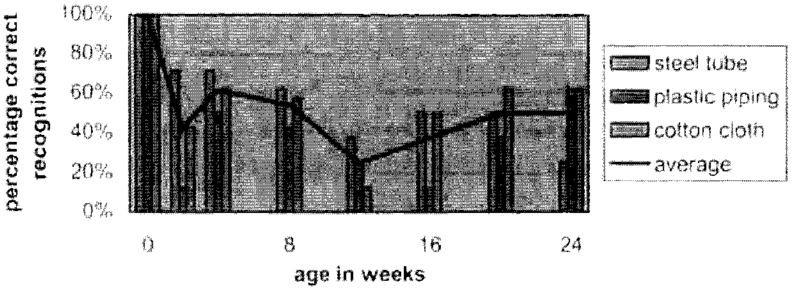

Fig. 2. Results of experiment 1: percentage correct recognitions of objects scented by all the people of different material type differing in age by dog A.

Table 2
Correct recognitions, no recognitions and incorrect choices based on pieces of cloth of increasing age by 10 dogs

	Age of cotton cloths (weeks)							
	0	2	4	8	12	16	20	24
Correct recognition	10	6	6	6	6	6	5	3
No recognition	0	2	3	3	1	2	4	6
Incorrect choice	0	2	0	1	1	1	1	0

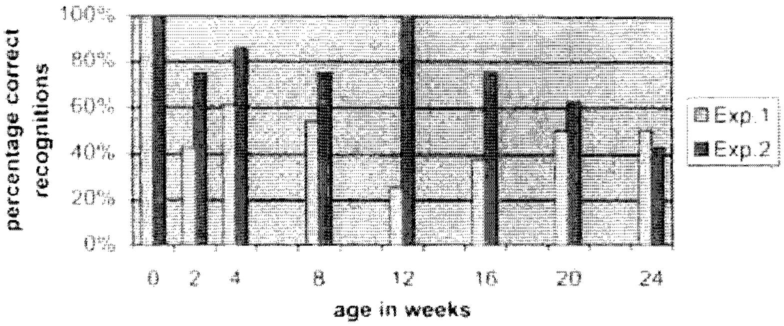

Fig. 3. Averaged results of experiment 1 (dog A) and experiment 2 (excluding the results of the two non-performers in experiment 2).

4. Discussion

The main objective of the experiments was to examine the effect of the ageing of odour evidence on the performance of the dogs in scent identification line-ups. Before evaluating the results of the experiments in this respect, a number of points that impact the performance of the dogs need to be discussed.

A first point is the effect of the set-up itself. The first experiment was designed as a "quick and dirty" set-up to collect a lot of information in a short period of time. A detailed examination of the protocol itself showed that it did not lead to systematic errors. The only "weak" point in the set-up was that the dog did not respond to the target odour if it was the first odour it came upon as readily as it did when it came to the target later in the circle. A major difference with the usual line-up protocol was the time it took to conduct. The circle-protocol took approximately 15 min, much longer than the usual protocol that only takes 4 min, or a training session, which usually does not last longer than 5 min. The performance dip at 12 weeks may well be an artefact of this prolonged working time. In terms of behaviour, a lack of interest was sometimes clear halfway through but with some additional playing the handler was able to revive the interest of his dog in the proceedings. The second set-up followed the current Dutch operational protocol. The difference here is that in operational cases, the handlers may decide which dogs they will use for a given case. In the experiment, participation was compulsory. The freedom to decide which dogs to deploy leads to a lower number of disqualifications in operational cases (less than 5%) than was found in these compulsory experimental line-ups (16.7%).

A second point that needs discussion is the effect of material type on the performance of the dogs. The results of the first experiment showed that there was no significant effect of material type. This is in line with earlier research [4], but seems to contradict the experience of training these dogs. Training experience shows that steel tubes are recognised most easily, followed by hard, smooth plastic, and cloth is recognised less easily. It may be that these subtle differences are overcome during the years of operational work. However, care should be taken when extrapolating from data presented here.

A third point is the possible effect of the scent source on the performance of a dog. It has been shown that occasionally a dog responds to or shows a marked interest in the odour of a particular person, and will do this in different contexts and at different times [5]. This observation is in line with training experience that dogs find the odour of one person much more difficult to recognise than another, and that on the other hand some odours are "easy". Experience with DNA profiles has also shown that some people are "shedders": they leave more body material behind than others, enabling forensic technicians to trace them more easily. Whether these "shedders" are also people whose scent is easily recognised by dogs is unknown. In the second experiment, two dogs did not perform well. This may have been caused by two poor scent-givers. However, both of these non-performers were retired early due to lack of consistent field performance. Three other dogs have since been retired for behavioural problems (not related to performance in scent identifications) or old age and five are still operational.

The general objective of the experiments was to determine the effect of ageing of articles collected at crime scenes on the performance of dogs in scent identification line-ups. Both the experiments show a marked drop from 100% recognitions with fresh material, but results after are lower and vary. Some of the factors that may have caused this variation have been discussed above, but such variability in performance has been described earlier and is not uncommon. The difference in the results between the two experiments described in this paper show, once again, the effect of experimental set-up on the performance of dogs in scent identification line-ups and illustrate how careful one must be in extrapolating data [6]. A number of other observations are worth noting.

First: training experience has shown that when training on aged objects, one needs to train the dog up to 5–6 day-old material and then it can recognise much older material. Secondly, recently a small-scale experiment with five Dutch and four German dogs was conducted. Here, material that had been scented by two persons more than 7 years earlier was used in scent identification line-ups. In total 12 line-ups were conducted, and 8 of these led to a positive identification [7]. This is very much in line with the results of the experiments described in this paper. And lastly, analysing data from Dutch operational cases leads to the observation that there seems to be little difference in the performance of dogs using material up to half a year-old, or using material that is between half a-year and 1-year-old.

In general, the conclusion can be drawn that after the initial drop, ageing does not seem to diminish the recognition of the dogs significantly. However, important questions are why this initial drop occurs, and how it can be prevented.

One answer to the cause of the initial drop may lie in the volatile nature of scent. Not all molecules evaporate at the same rate. Residue left on objects consists of different components. Even if initially the same residue is left on two objects but at a different time, the headspace above the objects at a later time can be different if the components of the residue have different vapour pressures. Since the objects are stored in glass jars, one can expect some kind of steady state to evolve since the amount of evaporation is limited, explaining the levelling off in the performance. Another explanation can be found in work that has been done in the quest to develop latent fingerprints. Latent fingerprints consist predominantly of fatty acids, and different laboratories have studied the composition of these prints [3]. Left out in the open air, it seems that large, unsaturated fatty acids are broken down into smaller, saturated fatty acids. The majority of these changes occur within the first week. This could also have a significant effect on the headspace, and depending on which components the dog is using as a cue, this could also explain the drop in performance.

Freezing the material, or storing it at low temperatures, could perhaps prevent the drop in performance. Scent pads used for bloodhound trailing in the United States are usually frozen, but there is no experimental basis supporting this choice. Further work is necessary in this field, and should also incorporate scent sampling methods and materials, and packaging.

Acknowledgements

I would like to thank the Dutch handlers who performed the experiments with their dogs for their work, and their organisations for allowing them to participate. I would also like to thank the German handlers and the ID-dog section of the Police Dog School of Nordrhein Westfalen police force in Germany for their participation.

References

[1] A. Schoon, R. Haak, K9 Suspect Discrimination, Detselig Enterprises Ltd., Canada, 2002.

[2] J. Wójcikiewicz, Scientific Evidence in Judicial Proceedings, Institute of Forensic Research Publishers, Krakow, 2000.

[3] R.S. Ramotowski, Composition of latent print residue, in: H.C. Lee, R.E. Gaensslen (Eds.), Advances in Fingerprint Technology, CRC Press, Boca Raton, 2001, pp. 63–104.

[4] G.A.A. Schoon, The performance of dogs in identifying humans by scent, Thesis, Leiden University, The Netherlands, 1997.

[5] G.A.A. Schoon, A first assessment of the reliability of an improved scent identification line-up, J. Forensic Sci. 43 (1) (1998) L70–L75

[6] G.A.A. Schoon, The effect of experimental set-up parameters of scent identification line-ups on their reliability, Problemy Kryminalistyki 236 (2001) 43–49.

[7] G.A.A. Schoon, Internal Report on the International Workshop for ID-Dog Handlers in Stukenbrock, 29 September–2 October 2003.

APPENDIX C

HOW SCIENCE WORKS
Evaluating Evidence in Biology and Medicine

Stephen H. Jenkins

OXFORD
UNIVERSITY PRESS
2004

OXFORD
UNIVERSITY PRESS

Oxford New York
Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Library of Congress Cataloging-in-Publication Data
Jenkins, Stephen H.
How science works : evaluating evidence in biology and medicine / Stephen H. Jenkins.
p. cm.
ISBN 0-19-515894-6; 0-19-515895-4 (pbk.)
1. Biology—Methodology. 2. Medicine—Methodology. I. Title.
QH324.J46 2004
570'.28—dc22 2003049407

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

Chapter 3

Can Police Dogs Identify Criminal Suspects by Smell?

Using Experiments to Test Hypotheses about Animal Behavior

We know from our everyday experience with pets that animals have different sensory abilities than humans. In many cases, it's obvious that these sensory abilities far exceed our own. Even the most casual observers, for example, would soon realize that their dogs live in a world dominated by odors and that they can detect and distinguish odors unknown to the owners. Indeed, some of the most fascinating stories in biology involve the discovery of specialized sensory abilities in particular types of animals that are totally lacking in humans. A classic example is Donald Griffin's (1986) discovery of echolocation by bats. Many bats emit streams of high-frequency sounds, well above the highest frequency that we can hear. These sounds bounce off objects in the environment, and the bats use the resulting pattern of echoes to navigate under pitch-black conditions at night. They also use echolocation to locate their prey, such as moths and other flying insects. This is an exquisite adaptation for success for a highly mobile, nocturnal, aerial predator of small prey that are also highly mobile.

The excitement of Griffin's discovery arises from the fact that, as humans, our primary tools for learning about the world are our own sensory abilities of taste, touch, smell, and especially hearing and vision. This has always been the case and is still largely true, despite the complex machinery for collecting and analyzing data that we associate with modern science (after all, the output of the machines generally has to be looked at or listened to by people in order to be interpreted). Because we depend on our limited sensory abilities for doing science, learning about sensory abilities of other animals that differ qualitatively or quantitatively from our own is particularly challenging. In the case of echolocation, European scientists discovered in the late 1700s that blinded bats could navigate in a room but bats whose ears were plugged with wax could not. Griffin repeated these experiments in the 1940s but was able to use new acoustic equipment to record the high-frequency sounds of the bats, finally solving the mystery of how bats navigate that had intrigued the European researchers 150 years earlier.

Donald Griffin (1986) describes his discovery of echolocation in a wonderful book for a general audience called *Listening in the Dark*. It's a tale of imagination, invention, and the design and execution of critical experiments. I'll use an example that is somewhat less exotic but especially well suited to illustrating the experimental method. It is the opposite of the echolocation story because it is about the *limits* of the olfactory abilities of dogs, which are often assumed to be virtually unlimited . Rather than inspiring our awe that animals can do something that we never imagined would be possible, such as navigate by echolocation, this story about the sense of smell in dogs shows how common knowledge can sometimes get ahead of scientific evidence, with significant practical consequences. In debunking the health benefits of vitamin C in Chapter 2 and the olfactory abilities of dogs in this chapter, I don't want to leave you with the impression that experiments are always used to discredit popular hypotheses. The next chapter will illustrate the positive role of experiments combined with other kinds of evidence to evaluate hypotheses.

THE MYTHIC INFALLIBILITY OF THE DOG

Humans have appreciated the olfactory abilities of dogs since antiquity and have developed various breeds to capitalize on these abilities. The use of dogs to assist in hunting is the oldest and probably most familiar example, but trained dogs are also used in various law enforcement tasks such as tracking fugitives and sniffing out narcotics. For example, James Earl Ray, who was convicted of killing Martin Luther King, Jr., escaped from the Brushy Mountain State Penitentiary in Tennessee but was tracked by two bloodhounds and quickly returned to prison. Dogs are also used by the police to identify suspects in lineups. Just as human witnesses to a crime may be asked to select a matching suspect from a lineup, a trained dog may be given a tool or piece of clothing from a crime scene to smell and then be presented with a lineup that includes a suspect and several other people. If the dog shows by its behavior that the odor of the suspect matches that of the object from the crime scene, testimony to this effect by the dog's handler is typically admissible in a trial. In fact, Andrew Taslitz reported in 1990 that thousands of these lineups have been done in the United States since 1920, contributing to convictions for robbery, rape, and murder and sentences up to life imprisonment or death.

What is the legal basis for admitting evidence from scent lineups in trials? According to Taslitz, this basis is remarkably weak, at least in the United States. Most courts use four criteria, which were developed for deciding the validity of tracking and simply transplanted to scent lineups: that the dogs should belong to a breed "characterized by acuteness of scent and the power to discriminate among individual human beings" (1990:120), that they be trained for tracking, that there be evidence of their reliability in tracking, and that there be other evidence independent of the dogs that is consistent with the scent identification. In many cases, convictions have been based primarily on scent identification by trained dogs, with additional evidence being only circumstantial. In practice, courts have relied almost exclusively on the claims of handlers about the abilities of their dogs. For example, in one robbery case, a handler testified that a tracking dog named Bobby was 100% accurate in training, as well as in four previous criminal cases. Both the original court and an appeals court accepted this testimony as "ample proof of reliability to justify admitting the results of Bobby's tracking" (1990:55).

Taslitz suggests that the faith of judges and juries in scent identification by dogs is rooted in a kind of mystical belief in man's best friend. Our culture has numerous legends, as well as true stories, attesting to the loyalty of dogs. In Homer's *Odyssey*, for example, Odysseus returns home after 20 years and is recognized by his aged dog Argus, who dies in the process of greeting his master. Argus probably recognized Odysseus after this long absence at least partly by smell. Because dogs are known for their loyalty, honesty, and integrity, we tend to accept evidence based on scent identification by dogs relatively uncritically. By contrast, imagine using *cats*, with their reputation for deviousness, in this way.

Because of "the mythic infallibility of the dog," in Taslitz's (1990:20) words, defense lawyers have so far been unsuccessful in persuading courts to apply the Frye Rule to evidence from tracking or scent lineups. The Frye Rule states that a scientific principle should be well established and generally accepted by scientists in the appropriate field for expert testimony based on that principle to be used as evidence in court. The purpose of the Frye Rule is to prevent juries from being swayed by the testimony of experts when the scientific foundation for that testimony is weak or nonexistent. In several cases, defense attorneys have argued that convictions should be overturned because the Frye Rule was not applied to the testimony of dog handlers. Judges have given various reasons for denying this argument, including the claim that such identification is not based on science, so the Frye Rule doesn't apply. This is clearly faulty logic. The olfactory abilities of dogs are subject to experimental testing, and a few such experiments have in fact been done (Brisbin et al. 2000). What is the evidence that trained dogs can recognize unique odors of individual people and use this ability to accurately identify subjects in lineups?[1]

In an interesting study with twins, Peter Hepper (1988) tested the hypothesis that humans have individual odors that can be recognized by dogs. Hepper was particularly interested in the role of genetic and environmental factors in causing people to have different odors. Therefore, he used three sets of twins in his experiments. The first set consisted of male fraternal twins that were 2 to 3 months old. These twins were genetically different but had a common environment because they were being raised in the same home. In particular, they ate the same foods, so any effect of diet on body odor should have been the same for both members of a pair. The second set consisted of male identical twins[2] that were 2 to 3 months old. These twins were not only genetically identical but also shared a common environment. The third set consisted of male identical twins that were between 34 and 50 years old, lived separately, and ate different foods. These twins were genetically identical but had different environments.

Four dogs were thoroughly trained in scent discrimination before being used in the experiments. This doesn't mean that the dogs learned *how* to distinguish between two similar scents; if they had an ability to discriminate, it was probably innate. Instead it means that the dogs were trained to show by their behavior that they were making a choice. The basic protocol was to wash each of the subjects with the same soap, then carefully rinse off the soap (presumably the adult subjects washed themselves). Four T-shirts for each pair of twins were also washed with the same detergent. Each twin then wore one T-shirt for 24 hours and a second T-shirt for the next 24 hours. Finally, a dog was presented with a T-shirt scented by one person, then given a choice between the second T-shirt worn by that person and one of the T-shirts worn by his twin. The dog's handler did not know which was the correct match.

The results of this experiment suggest that dogs can use *either* genetic or environmental factors to discriminate between people. The average percentage of correct choices was 89% for the infant fraternal twins sharing a common environment, 49% for the infant identical twins sharing a common environment, and 84% for the adult identical twins living in different environments. Since the dogs were choosing between two T-shirts in each trial, they should have been able to get 50% of their choices correct just by chance. The 49% rate of success at distinguishing twins with the same genes and same environment is not different from chance performance, suggesting that the dogs got no useful cues for discrimination in this situation. This provides a baseline for comparison with the other two types of trials, where performance of the dogs was significantly better than chance. For the trials with adult identical twins, environmental differences such as diet must have been the basis for discrimination by the dogs. For the trials with infant fraternal twins, genetic differences must have been the basis for discrimination because infant identical twins could not be distinguished.[3]

Are these results sufficient to validate the use of dogs to identify suspects in scent lineups? Not by a long shot, for at least four reasons (not including the fact that 2-month-olds rarely commit crimes). First and least important, Hepper tested only four dogs, so we can't say if their abilities are common or unusual among dogs in general (this is least important because Hepper did show that *some* dogs have remarkable abilities to discriminate human odors). Second, the success rates were impressive, but the performance of the dogs was not perfect even after intensive training. Dogs made mistakes in identification 11% of the time in trials with infant fraternal twins and 16% of the time in trials with adult identical twins. Error rates of this magnitude that resulted in convictions of innocent people would be unacceptable. Third, dogs used in forensics often have to compare an odor from one body part of a perpetrator (e.g., head odor on a hat left at the crime scene) with an odor from another body part of a suspect in a lineup (the suspect and several control individuals use their hands to apply scent to a test object such as a metal cylinder). In Hepper's experiments, odors all came from the torsos of the twins. Just as individuals may differ in smell, various body parts of the same individual may also differ. Can dogs be trained to ignore this anatomical variation and detect a component of the olfactory signal that's common to all body parts of an individual, if such a common signal exists? Finally, Hepper's experiment differs from real-world forensic practice in that a suspect in a lineup may not be the same as the perpetrator of a crime. If this is the case, a dog given an object from a crime scene to sniff should pick no one from the lineup because no odors of these people would match the smell of the object from the crime scene. In Hepper's experiment, by contrast, one of the T-shirts presented to the dog was always a correct choice.

I. Lehr Brisbin and Steven Austad (1991) did a small experiment to see how the ability of dogs to distinguish between the scents of two different individuals was affected by the body parts that supplied the scents. They used three dogs and modeled their procedure after that used in competitions authorized by the American Kennel Club. In these competitions, dogs are required to select metal and leather dumbbells scented by their handlers' hands when given a choice between these and dumbbells scented by the judge's hand. This differs from forensic practice in an important way because we might expect that it would be easier for dogs to distinguish between the familiar odors of their handlers and the unfamiliar odors of judges than between the odors of various people in lineups, which would all be equally unfamiliar. This means that the experiment of Brisbin and Austad gives the benefit of the doubt to the hypothesis that dogs can generalize across body parts to identify individuals. If their dogs were able to do this, it might be because the handlers' odors were so familiar, which would provide only weak evidence that dogs could generalize in scent lineups. However, if the dogs used by Brisbin and Austad were not able to generalize, it seems unlikely that dogs could do so in the more challenging circumstances of the lineups.

The dogs were trained by using standard guidelines of the American Kennel Club for obedience training of "utility dogs" for at least 6 months, then further trained in the specific procedures used by Brisbin and Austad in their experiments. In each trial, two scented dumbbells were placed on a board about 10 feet from the dog and handler while the dog and handler were facing in the opposite direction. When the dumbbells were in place, the dog and handler were instructed to turn around to face the board. The handler gave the dog a command to go to the board. Sniffs of each of the dumbbells were recorded, as well as which dumbbell was retrieved. The experimenters positioned the dumbbells in such a way that the handler didn't know which was the correct choice.

The dumbbells were scented by having a person hold a dumbbell in his or her hand for 30 seconds or using tongs to position the dumbbell in the interior crook of the elbow and having the person hold it tightly there for 30 seconds. Several types of trials were conducted. First, the dogs had to choose between (1) a dumbbell scented by the handler's hand and one with no human scent or (2) a dumbbell scented by the handler's elbow and one with no human scent. The dogs had no trouble with these discriminations, averaging 93% correct in the first case and 86% correct in the second case. The dogs also had little trouble in the second type of trial, distinguishing a dumbbell scented by the handler's hand from one scented by a stranger's hand, which is comparable to the task used in obedience competitions. Success rates of the three dogs were 69%, 70%, and 90%, for an average of 76%. However, when given a choice between a dumbbell scented by their handler's elbow versus one scented by a stranger's hand, the dogs were less successful. Success rates were 70%, 57%, and 46%, and the average of 58% was not significantly different from the chance performance of 50%. Since the dogs were trained to identify the hand odors of their handlers, they may have been confused by differences between elbow odors of their handlers and hand odors of strangers. It is interesting that they could distinguish elbow and hand odors of their handlers, with an average success rate of 78%.

In a nutshell, these results suggest that different body parts of the same person have different odors. This shouldn't be surprising, although the fact that the elbow and hand of the same arm smell differently may be a bit curious. More important, the results suggest that dogs do not automatically generalize from one body part of an individual to another to discriminate between two people with individually distinctive odors, *if the dogs are trained by using standard methods*. It's quite conceivable that a training regime could be devised to improve the performance of dogs in this task, but present methods that are used not only to train dogs for competitions sponsored by the American Kennel Club but also for scent identification in police work are not adequate. This undermines one of the fundamental assumptions of the use of evidence from scent lineups in court.

A group of researchers led by Ray Settle of the Police Dog Training School in Preston, England, developed and tested a training routine that they thought might be more effective than standard methods of training dogs to generalize across body parts in identifying individuals (Settle et al. 1994). Settle's group used seven dogs of various breeds. They collected body scents from more than 700 volunteers from various schools, a local business, and a nursing home. Each volunteer was given four pieces of cotton cloth that had been washed and placed in a glass jar. The volunteer was asked to place each piece of cloth on a different part of his or her body for 30 minutes, then replace the scented cloths in the jar. The choice of body parts to be scented was up to the volunteers.

The dogs were tested by giving them one cloth from one of the volunteers to sniff and then either presenting them with a group containing another cloth scented by the same volunteer plus five cloths scented by five other volunteers or presenting them with six steel tubes that had been held by volunteers for 5 minutes. These procedures are similar to those used in actual scent lineups, at least in Europe. Handlers trained their dogs in a series of progressively more difficult discriminations. For example, in an early step in training, the dogs had to identify a scented cloth when it was placed in the training room with five others that had been washed and handled with tongs, so they had no human scent. In successive steps, the dogs had to discriminate a target scent from one other human scent, then two others, and so on. Training lasted for 9 months.

After this regimen, the dogs were correct 80% of the time on average in the first type of test, in which they had to discriminate among six cloths scented by different people, and 85% of the time in the second type of test, in which they had to discriminate among six steel tubes handled for 5 minutes by different people. In each case, since one of the six choices matched the target scent, we would expect the dogs to be correct one-sixth of the time (17%) purely by chance. Accuracy of 80 to 85% is much better than chance performance, implying that the dogs really had learned to distinguish individual human odors in a situation similar to that used in actual police lineups. Furthermore, Settle and his colleagues suggested that the dogs were identifying individuals regardless of whether or not the body part used as a source of the target scent was the same as that used in the lineups.

These results of Settle's group are inconclusive, however, because the volunteers who provided cloths scented by four different parts of their bodies handled all of the cloths and placed them together in a closed glass container that was returned to the experimenters. In handling the cloths, hand odors may have been transferred to them; in keeping them in closed containers for up to 4 days, odors may have been transferred among them. This means that the odors on four cloths prepared by the same volunteer were probably more similar to each other than, for example, the odors on the dumbbells held in the hand and the elbow by a subject in the experiment of Brisbin and Austad (1991). Because of this likely odor contamination in the study by Settle and his colleagues, it may have been easier for the dogs to match a target scent to one of the scents in the lineup than if it had come from a distinctly different part of the body than the part used to create the lineup. Therefore, this study doesn't restore much faith in the validity of evidence from scent lineups, because it doesn't convincingly show that dogs can generalize from one body part to another to distinguish odors from different people, even though the training used by Settle's group was more extensive than that normally given to police dogs.

The most extensive and most recent set of experiments dealing with scent identification by dogs in police lineups was done by Gertrud Schoon for her dissertation at the University of Leiden in the Netherlands.[4] She is affiliated with the Department of Criminalistics and Forensic Science and the Ethology Group of the Institute of Evolutionary and Ecological Sciences at the university, indicating that she has broad training and diverse interests. Real progress in science often comes from people who bring a new perspective to a long-standing problem. In this case, Schoon's background in ethology (animal behavior) may have helped her to find creative ways of dealing with some of the practical problems of forensic science that we've been discussing.

Schoon (1996, 1997, 1998) examined several aspects of scent identification to find ways to improve the training of dogs and the operation of scent lineups, but I will discuss just one part of her work that addresses one of the most important pitfalls of these lineups. This is the possibility that a dog will falsely accuse a suspect by selecting the suspect's scent from a lineup when this scent does not match the target scent from an object left at a crime scene. Unlike all the experiments described so far that used a *match-to-sample* design, in which an odor matching the sample was always present in the array with which the dogs were tested, in the complicated real world of police investigations innocent suspects are sometimes arrested. In this case, the correct choice of a dog would be to select none of the odors in a lineup because none would match that from the crime scene. But there are various factors that might work against dogs making this response. They might select the odor that was *closest* to the odor on the target object, even if it wasn't identical. In fact, when human witnesses to crimes are shown pictures of several potential subjects, they tend to pick one that looks most similar to the person they saw at a crime scene, even if that person was not actually there (Wells et al. 2000). In scent lineups, the control scents often come from police officers who may be familiar to the dog, so the dog may pick the odor from the lineup that is least familiar, regardless of whether it matches the scent of the target object. The handler may believe the suspect is guilty and therefore reward the dog for making any selection at all. If the lineup consists of a suspect and several control individuals, none of whom are known to the handler, the handler may pick out the likely suspect by using visual cues and communicate this identification to the dog unconsciously.[3] Even if the lineup consists of objects like metal cylinders that were scented by using standardized methods, the handler, who is likely to be a police officer familiar with the case, may believe the suspect is guilty and therefore unconsciously encourage the dog to choose one of the cylinders when the correct response if the suspect was really innocent would be no choice. To date, Schoon is the only researcher who has rigorously tested the possibility of false identification of suspects based on dogs and scent lineups.

Schoon did her experiments with six tracking dogs that were trained for police work in the Netherlands. Each dog was used in 10 trials. In each one, a target scent was prepared by having a volunteer police employee treat an object as if it had been found at a crime scene. These objects were screwdrivers, wrenches, pistol buttplates, sweatshirt cuffs, and scent samples from the seat of the volunteer's car. In preparing these target scents, the volunteers were acting as if they were the perpetrators of a crime.

To prepare odors for scent lineups, Schoon gave volunteers two glass jars with six stainless steel cylinders in each. The volunteers were instructed to handle the cylinders in each jar for 5 minutes. In addition, one of the volunteers handled a piece of polyvinyl chloride (PVC) tubing, providing a "check" scent that was used as described below to be sure the dogs didn't have an inherent preference for one of the odors in the lineup but were really making a choice based on matching an odor in the lineup to a target odor. For each trial with a dog, one volunteer was designated as a suspect. In half of the trials with each dog, this "suspect" was the same as the "perpetrator" who had scented the objects from a hypothetical crime scene; in the remaining trials, the suspect and perpetrator were different people.

Each trial consisted of four tests of discrimination in which seven scented cylinders were laid out for a dog. One cylinder had the scent of the suspect, one cylinder had the scent of the person who provided the check scent, and the other five cylinders had scents of five other volunteers that Schoon called decoys. In the first two trials, the dog was allowed to smell the check scent on the PVC tubing, then shown the lineup of seven cylinders. In this case, the correct choice was to select the cylinder scented by the person who made the check scent. If the dog made the wrong choice in either of these tests, it was disqualified for that trial. For example, the dog might be disinterested in working that day and make no choice, or it might select the suspect's scent, indicating a preference for that scent even though it didn't match the check scent. In these cases, Schoon believed it was invalid to test the dog's ability to compare a perpetrator's and suspect's scent. One of her pragmatic suggestions is that police departments begin using this kind of performance check in actual lineups.

After these two trials, the check scent was removed from the lineup and the dog was given the scented object made by the person pretending to be the perpetrator. Then it was given two tests with six scented cylinders—one scented by the simulated suspect and five scented by decoys. Recall that in half of the trials with each of the six dogs, the suspect was the same as the perpetrator, whereas in the other half the suspect and perpetrator were different, and the handler did not know which trial was which. Since there were five types of objects, there were 10 trials per dog, for a total of 60 trials.

The performance of the dogs in this experiment was not particularly impressive. In 30 of the 60 trials, the dogs were disqualified because they made errors in the check tests. In the remaining trials in which the suspect was the same as the perpetrator, the dogs correctly selected the scent of the suspect in four trials, they selected a decoy's scent in five trials, and they made no selection in two trials, for a success rate of 4/11, or 36%. In the trials in which the suspect was not the same as the perpetrator, the dogs correctly made no selection in nine cases, they selected the suspect's scent in one case, and they selected a decoy's scent in nine cases, for a success rate of 9/19, or 47%. It's interesting that the dogs made some choice in a majority of these trials, suggesting that it may have been difficult for them to resist picking one item from the lineup even when there was no match to the scent on the object from the hypothetical crime scene. However, the effective error rate was only 1/19, or 5%, because in nine cases the dogs selected a decoy, who would presumably be known to be innocent in a real lineup. These estimates assume that the conditions of this experiment with simulated suspects and perpetrators are representative of actual forensic practice. Also, the estimates are not very precise because they are based on a small number of successful trials with only six dogs.

INTERPRETING EVIDENCE

Ignoring the limitations of its small sample size, what can we learn from this study? There are two categories of errors that can occur in methods used to match suspects to evidence from a crime scene, whether this evidence be fingerprints, tissue samples containing DNA, eyewitness accounts, or odors on objects identified by dogs. These errors are false positive identification, in which an innocent suspect is convicted because of an incorrect match between the evidence and the suspect, and false negative identification, in which a guilty suspect is acquitted because of failure to make a match between the evidence and the suspect. The general principle of jurisprudence, that a person is considered innocent until proven guilty beyond a reasonable doubt, implies that the first type of error is considered more serious than the second, at least in contemporary Western society. McCauliff (1982) surveyed 171 judges and found that the most common interpretation of "beyond a reasonable doubt" was that the chance of false positive identification was 10%. This means that innocent suspects might be convicted as frequently as 10% of the time, or, said another way, that conviction is reasonable if the likelihood of guilt is greater than 90%. Of course, in our legal system of trial by jury, the ultimate determinant of the meaning of "beyond a reasonable doubt" is the collective opinion of the jurors.

Before applying these ideas to the problem of scent identification, I'd like to discuss a seemingly different but in fact perfectly parallel problem, which is easier to analyze because fewer assumptions are necessary. Consider a diagnostic test for a disease, such as the occult-stool test for colorectal cancer in which blood in the stool is used as an indication that a person might have cancer in the lower portion of the digestive tract (Hoffrage et al. 2000). We can summarize the values needed for our calculations in a 2 × 2 table in which the two columns represent people with and without the disease and the two rows represent positive and negative results of the test (Table 3.1). The sensitivity of the test is 50%; that is, if a person has the disease, the chance of a positive test result is 50%. This means that the occult-stool test misses 50% of the cases of colorectal cancer. These are false negatives, represented by the value of 0.5 in the lower-left cell of the table.

The upper-right cell of the table shows the probability of a false positive test result: an occult-stool reading that indicates colorectal cancer in a person who does not actually have the disease. This probability is only 3%. In other words, for a person who is not afflicted with colorectal cancer, the chances are 97% that the occult-stool test would correctly be negative. But the most important issue is how to interpret a positive test result for someone whose

Table 3.1. The occult-stool test for colorectal cancer.

		Disease Status of Individual	
		Has Colorectal Cancer	Does Not Have Colorectal Cancer
Results of Occult-Stool Test	Positive	0.5	0.03
	Negative	0.5	0.97

Values in the first column are probabilities of positive and negative test results for people with colorectal cancer; values in the second column are probabilities of positive and negative test results for people who do not have colorectal cancer (Hoffrage et al. 2000).

disease status is not known. Stated more personally, if your doctor did an occult-stool test and reported a positive result, what are the chances that you have colorectal cancer?

Based on the top row of Table 3.1, you might assume that your chances of having colorectal cancer in this case are pretty high because the probability of a positive test result in a person with colorectal cancer is much larger than the probability of a false positive (0.5 versus 0.03). However, the significance of this depends on the frequency of colorectal cancer in the general population, which is only about 0.3% (Hoffrage et al. 2000). Why is this important? The easiest way to see how this information affects the analysis is to imagine applying the probabilities in Table 3.1 to a large, hypothetical group of people. Suppose we consider a group of 10,000. If this group is representative of the general population, 30 would have colorectal cancer. Of these 30, 15 would have a positive occult-stool test and 15 would have a negative test, based on the first column of Table 3.1. Of the remaining 9,970 people who do not have colorectal cancer, we expect there to be about 299 who would have false positive tests (0.03 × 9970). These numbers are shown in Table 3.2. Looking at the first row of Table 3.2, we see that the expected total number

Table 3.2. Test results and disease status of 10,000 hypothetical individuals, where the frequency of colorectal cancer is 0.3%.

		Disease Status of Individuals		
		Have Colorectal Cancer	Do Not Have Colorectal Cancer	Totals
Results of Occult-Stool Test	Positive	15	299	314
	Negative	15	9,671	9,686
Totals		30	9,970	10,000

I constructed this table by first subdividing the total of 10,000 individuals into 30 with colorectal cancer and 9,970 without colorectal cancer because 0.3% of 10,000 = 30 and 10,000 − 30 = 9,970. For those with colorectal cancer, half would be expected to have a positive occult-stool test and half would be expected to have a negative test based on Table 3.1, producing the values of 15 in the first numerical column here. For those without colorectal cancer, 3% would be expected to have a positive occult-stool test based on the upper-right cell in Table 3.1, and 3% of 9,970 = 299. Summing across the first row gives 314 positive tests among 10,000 people. Only 15 of these positive tests (5% of 314) would occur in people who actually have colorectal cancer.

of positive test results is 314, of which only 15 are actually associated with cancer. That is, only about 5% of people with a positive occult-stool test would actually have colorectal cancer, and you might be fortunate enough to be in the 95% without cancer, despite a positive test. Even though the likelihood of a false positive result for an individual without colorectal cancer is relatively small (3%) compared to the likelihood of a correct positive result for an individual with cancer (50%), the number of people without cancer is so much larger than the number with cancer that most positive test results occur in people without cancer.

Now let's return to scent identification by dogs. The results of Schoon's research are summarized in Table 3.3, which has the same format as Table 3.1. The critical question is similar to that for diagnostic testing for colorectal cancer: if a dog identifies a suspect as having an odor matching that from a crime scene, what is the probability that the suspect is guilty? Just as in the cancer example, we need more information than the values shown in Table 3.3 to answer this question. In the cancer example, the additional information was the frequency of colorectal cancer in the population. The parallel information for the scent-identification example would be the number of potential suspects for the crime. If we assume that only one person committed the crime and that there are 10 possible suspects, the proportion of suspects that is guilty is 1/10, or 10%, just as the proportion of people with colorectal cancer is 30/10,000, or 0.3%. Unfortunately, for most crimes it's not very clear how many potential suspects there are. However, imagine one of those classic murder mysteries on an estate in the English countryside. The owner, an eccentric bachelor, has 10 servants (the gardener, the butler, the cook, etc.). He invites 10 guests for a weekend of hunting. On Saturday evening, the owner is discovered murdered. A handkerchief with no identifying marks has been left on the floor of the library where the owner's body is found. The local constable brings his trained dog to match the scent of the handkerchief to that of one of the suspects. It's obvious that one of the 10 servants or 10 guests committed the dastardly deed, so the calculations can be made just as in the cancer example (to keep things reasonably simple, we'll assume that

Table 3.3. Scent identification by dogs in lineups based on Gertrud Schoon's (1998) research.

		Status of Suspect	
		Guilty (Suspect = Perpetrator)	Innocent (Suspect ≠ Perpetrator)
Identification by Dog	Positive (suspect selected)	0.36	0.05
	Negative (suspect not selected)	0.64	0.95

Values represent the proportion of trials in which the dogs made correct choices (the upper-left and lower-right cells), in which they failed to identify a guilty suspect (false negatives in the lower-left cell), and in which they incorrectly identified an innocent suspect (false positives in the upper-right cell).

there was only one murderer). These calculations are shown in Table 3.4. The bottom row shows the 20 total suspects divided into two groups, one guilty person and 19 innocent ones. The first column of Table 3.3 remains the same in Table 3.4 because we assume there is only one guilty person. But the values in the second column of Table 3.3 are increased in Table 3.4 because there are 19 innocent suspects who might be falsely identified by the dog. Specifically, the values in the second column of Table 3.3 are multiplied by 19 to get the second column of Table 3.4. This implies that if the dog selects a suspect (the top row of Table 3.4), the chance that the suspect is guilty is only 27% (0.36/1.21 = 0.27). Even though the probability of a false positive (0.05) is much lower than the probability of correctly picking the guilty person (0.36), there are so many more opportunities for the dog to select an innocent suspect than the guilty person (19 versus 1) that the likelihood that it will pick an innocent person is 73%.

Most criminal cases in which a scent lineup might be used aren't as straightforward as this because the total number of possible suspects isn't known. However, there is often other evidence that links a particular suspect to a crime. At least in theory, this other evidence can be used to estimate the likelihood that the suspect is guilty, independently of whether a dog selects the suspect's scent from a lineup. This is called the *prior probability* of guilt; that is, the probability of guilt before taking into account the results of the scent lineup. In the example of the English murder mystery, the prior probability of guilt is 1/20, or 5%, for each suspect because there are exactly 20 suspects and we have no other evidence pointing toward any one of the 20. This prior probability is represented in the last row of Table 3.4. In the scent lineup, the constable's dog matches the scent of one of the suspects to the odor on the handkerchief left by the victim's body. Because trained dogs really do have some ability to identify individual humans by smell, this increases the likelihood that the suspect identified by the dog is guilty. After the lineup, the so-

Table 3.4. An application of Schoon's (1998) results on accuracy of scent identification by dogs.

		Status of Suspect		
		Guilty (Suspect = Perpetrator)	Innocent (Suspect ≠ Perpetrator)	Totals
Identification by Dog	Positive (suspect selected)	0.36 × 1 = 0.36	0.05 × 19 = 0.95	1.31
	Negative (suspect not selected)	0.64 × 1 = 0.64	0.95 × 19 = 18.05	18.69
Totals		1	19	20

We assume that there are 20 possible suspects, 10 servants and 10 guests, only one of whom is guilty of murdering the owner of an English country estate. This provides the totals in the bottom row. The totals for each column are multiplied by the appropriate probabilities in each cell of Table 3.3 to get the values in this table, which can be used to compute the probability that an innocent suspect is mistakenly identified as guilty. This probability is 0.95/1.31 = 0.73.

called *posterior probability* of guilt is 27%. This is a substantial increase over the 10% prior probability of guilt, but it is far short of the standard expressed by the phrase "beyond a reasonable doubt."[6]

Suppose we wanted to be 90% sure of blaming the correct person for the crime. How many suspects would we have to exclude based on other evidence so that the posterior probability of guilt following scent identification by the dog was greater than 90%? The best we can do is to reduce the pool of suspects from the original 20 to two, based on other evidence. In this case, after the dog picks one of these suspects, the posterior probability that this suspect is guilty is 88%: not quite reaching the standard for establishing guilt beyond a reasonable doubt.[7]

In practice, other evidence besides scent identification is often qualitative and not easily converted into the prior probabilities used in this example. However, if Schoon's results on the abilities of trained dogs to match human scents are accurate, the method of scent lineups doesn't seem to have much credibility. Even in the ideal and unlikely situation that other evidence reduces the pool of potential suspects to two, there is still about a 12% chance that the dog would pick the wrong suspect from a lineup.

Let's consider a parallel situation in which these kinds of calculations are helpful but actually validate a common forensic method rather than casting doubt on it. This is the well-known use of DNA to match suspects to blood or tissue samples found at a crime scene (Gomulkiewicz and Slade 1997). We can set up tables just like Tables 3.3 and 3.4 to show the calculations. With current technology, the likelihood of missing a match between a truly guilty suspect and a sample of that suspect's DNA from a crime scene is very low, certainly less than 0.5% and perhaps in principle equal to zero. This is a false negative or false mismatch, shown in the lower-left cell of Table 3.5. The probability of a false positive depends on four factors: the possibility that an innocent person shares the same DNA profile as the perpetrator of the crime for the regions of DNA that were analyzed, the possibility of laboratory error such as contamination of a sample, the possibility that the innocent person left his or her DNA at the crime scene but did not commit the crime, and the possibility that a DNA sample from the innocent person was planted at the crime scene. If the latter three possibilities can be ruled out, the probability of a false positive ranges from one in 100,000 to one in 1 billion, *for suspects who are not relatives of the perpetrator of the crime.*[8] Because brothers, for example, share 50% of their DNA, the likelihood of a false positive is as high as 0.26 for an innocent suspect who is the brother of the actual criminal. For Table 3.5, I assume that the pool of potential suspects includes only unrelated people, and I use an intermediate value for the probability of a false positive identification of one in 10 million.

Applying these values to the murder at the English country estate (and assuming that the mysterious handkerchief has blood on it that does not match that of the victim), Table 3.6 shows that the likelihood of guilt of a suspect whose DNA profile matches that of the blood on the handkerchief is

Table 3.5. Identification of suspects based on blood or tissue samples containing DNA collected at a crime scene.

		Status of Suspect	
		Guilty (Suspect = Perpetrator)	Innocent (Suspect ≠ Perpetrator)
DNA Collected at Crime Scene	Matches Suspect	0.995	0.0000001
	Does Not Match Suspect	0.005	0.9999999

The values in the table are the probabilities that DNA collected at the crime scene (1) matches that of the suspect if the suspect is guilty (the upper-left cell), (2) matches that of the suspect if the suspect is innocent (a false positive result in the upper-right cell), (3) does not match that of the suspect even though the suspect is guilty (a false negative result in the lower-left cell), and (4) does not match that of the suspect if the suspect is innocent (the lower-right cell). These values were derived from a review of the use of DNA evidence in court by Gomulkiewicz and Slade (1997).

0.995/0.9950019, which is greater than 99.99%. Even if the pool of potential suspects was much larger than 20, DNA evidence may be quite persuasive, provided factors like sloppiness in lab techniques or planting
crime scene can be excluded.

However, there may be situations in which even DNA evidence is not as conclusive as might be assumed. Suppose the *only* evidence available is DNA from a crime scene. The FBI and other law enforcement agencies have databases of DNA profiles for large numbers of individuals who have had various encounters with the legal system. The sizes of these databases are increasing daily. If the authorities have no other evidence, they may scan the database to see if there is a profile that matches that of the DNA from the crime scene. If there are 5 million profiles in the database and we assume that the guilty person is one of those 5 million, we would substitute 5 million for the overall total in the bottom right cell of Table 3.6. In this case, with one guilty person

Table 3.6. An application of DNA identification to a murder in an English country estate.

		Status of Suspect		
		Guilty (Suspect = Perpetrator)	Innocent (Suspect ≠ Perpetrator)	Totals
DNA Collected at Crime Scene	Matches Suspect	0.995	0.0000019	0.9950019
	Does Not Match Suspect	0.005	0.9999981	1.0049981
Totals		1	19	20

As in Table 3.4, we assume that there are 20 possible suspects, only one of whom is guilty. This provides the totals in the bottom row. The totals for each column are multiplied by the probabilities in the appropriate cell of Table 3.5 to get the values in this table, which can be used to compute the probability that an innocent suspect is mistakenly identified as guilty based on DNA evidence. This probability is 0.0000019/0.9950019, which is much less than 1%, if the assumptions about the use of DNA evidence discussed in the text are correct.

and 4,999,999 innocent suspects, the upper-right cell of the table would become 0.0000001 × 4,999,999 = 0.5. Therefore the top row of the table would be changed to 0.995, 0.5, and 1.445. Given a match between DNA from a crime scene and DNA from one of the 5 million individuals in the database, the probability of guilt for that particular individual would be 0.995/1.445, or 66%. Just as in the example of the occult-stool test for colorectal cancer, in which such a small proportion of the population has the disease that most false positive tests occur in healthy people, there is a substantial likelihood of a false positive result if DNA from a large number of innocent people is screened to find a match to DNA collected at a crime scene.

DNA evidence is certainly a powerful tool in forensic identification, but it can be misused in several ways, including the kind of fishing expedition described in the last paragraph (Roeder 1994). Although the calculations often depend on assumptions that can't be verified (e.g., that the total number of possible suspects is known), this is a valuable exercise because it introduces a systematic way of thinking about the credibility of evidence.

In this chapter we've examined several experiments dealing with the olfactory abilities of dogs in the context of forensic work. These experiments were not ideal in design and execution, but perhaps they are more valuable for learning some of the basic elements of experimentation because of their flaws than more rigorous experiments would have been. It might seem that it would be easy to study the behavior of dogs experimentally because they are so much more familiar to us than other animals, but in fact studying trained dogs means working with their handlers as well, and this can complicate experiments significantly.

Although the limited experimentation that has been done to date, mostly by Gertrud Schoon, casts doubt on the validity of scent lineups in "proving" that a suspect committed a crime, some of the other work that trained dogs do is still credible. This includes finding people buried by avalanches or collapsed buildings, sniffing out hidden narcotics, and tracking fugitives. Also, dogs clearly have the sensory ability to identify odors specific to individuals and thus determine if a suspect in a scent lineup has the same odor as an item from a crime scene, although not infallibly. The key is to devise a training method that enhances the accuracy and reliability of this identification.

Finally, this chapter illustrates a quantitative method for evaluating evidence. I used examples ranging from diagnostic tests for disease to matching DNA from a crime scene to DNA of a suspect, but this method has even broader applicability to hypothesis testing in general. In many cases, we may not have all the information necessary to apply this method. However, using this approach as a framework for thinking about the meaning of evidence can help us develop an appropriate level of skepticism about how scientific information is used in practical matters, as well as in basic research. The results of this approach are sometimes surprising, as in some of the examples we discussed. If jurors were better educated about these ideas, we would all benefit from more rational judicial proceedings.

RECOMMENDED READINGS

Gawande, A. 2001. Under suspicion: The fugitive science of criminal justice. *The New Yorker,* 8 January 2001, p. 50. Gawande briefly reviews attempts to test experimentally some basic assumptions of forensics, focusing on the veracity of eyewitness accounts of crimes.

Hoffrage, U., S. Lindsey, R. Hertwig, and G. Gigerenzer. 2000. Communicating statistical information. *Science* 290:2261–2262. This article discusses why many people have difficulty in interpreting information about probabilities, using the occult-stool test for colorectal cancer and other medical examples as illustrations.

Malakoff, D. 1999. Bayes offers a 'new' way to make sense of numbers. *Science* 286:1460–1464. The calculations in the last section of this chapter are based on Bayesian statistics, and Malakoff provides an introduction to this topic.

Taslitz, A. E. 1990. Does the cold nose know? The unscientific myth of the dog scent lineup. *Hastings Law Journal* 42:15–134. Taslitz reviews the use of dogs in legal proceedings.

8. The subjects in the initial trial were mostly white females with an average age of 36 and no other diseases. The FDA Advisory Committee wanted to see results for a broader range of subjects and more information on side effects before recommending approval. The committee also noted that pleconaril had to be started within the first 24 hours of onset of a cold to be effective and wondered about the practicality of this treatment (Food and Drug Administration 2002; Senior 2002).

Chapter 3

1. In 1993 the U.S. Supreme Court decided that the Frye Rule, which was promulgated in 1923, was superseded by the Federal Rules of Evidence, which became the legal basis for admissibility of evidence in federal cases in 1975. These rules are used in many state courts, as well as by federal judges. In its 1993 decision, the Supreme Court also specified four criteria for evaluating scientific evidence in court: (1) the methods underlying the evidence should be testable, (2) published or reviewed by scientific peers, and (3) generally accepted by the relevant community of scientists, and (4) the likelihood of error in the methods should be able to be estimated (Annas 1994). To my knowledge, this Supreme Court decision, *Daubert v. Merrell Dow Pharmaceuticals*, has not been applied to the use of dogs for scent identification in police lineups.

2. See Chapter 6 for a discussion of more precise terminology for classifying types of twins.

3. In a recent study, Wells and Hepper (2000) tested the ability of *humans* to distinguish between the scents of different *dogs*. They found that dog owners could pick out a blanket scented by their own dog when it was paired with a blanket scented by an unfamiliar dog, although they expressed no consistent preference between the two.

4. Schoon published her dissertation research in articles in *Applied Animal Behaviour Science* (1996), *Behaviour* (1997), and the *Journal of Forensic Science* (1998).

5. A horse named Clever Hans is a famous example of the ability of animals to attend to very inconspicuous movements of humans. Clever Hans, who lived in Germany in the early 1900s, purportedly could solve mathematical and other problems and indicate correct answers by tapping his hoof. Clever Hans was very popular with the public, especially after a group of scientists saw him demonstrate his talents and declared that there was no evidence of fraud on the part of his owner. This inspired an experimental psychologist named Oskar Pfungst to test Clever Hans in various ways, including the first double-blind experiments in psychology. If the owner didn't know the answer to a question or if a curtain was placed between the owner and Clever Hans, the horse never got the right answer. After long and careful observation, Pfungst finally discovered that the horse noticed barely perceptible and unconscious head movements or postural changes of the owner when the horse had reached the correct number of taps, and he stopped tapping at that point (J. L. Gould 1982).

6. Here is a similar problem to test your understanding of these calculations. Suppose there are two taxi companies in Reno, Blue Cab and Green Cab. Eighty-five percent of the cabs in town are blue while 15% are green. A cab was responsible for a hit-and-run accident at night; an eyewitness saw the accident and reported that the cab was green. The police tested the ability of the witness to determine the color of cabs at night and found that the witness was accurate 80% of the time. What is the probability that the cab involved in the accident was actually green?

7. Substitute one guilty suspect, one innocent suspect, and two total suspects in the bottom row of Table 3.4; then recalculate the top row of this table by multiplying

0.36 by 1 and 0.05 by 1 and adding these two values to get 0.41. Therefore, the probability that the suspect identified by the dog is actually guilty is 0.36/0.41 = 0.88.

8. Why are these probabilities of a false positive result so low? Laboratories that test DNA typically analyze several different regions of the DNA extracted from a sample of blood or other tissue. For each of these regions, several different sequences of DNA typically exist in a population. Two unrelated individuals might have the same sequence at some of these regions, but the probability that they will have the same sequence at all regions tested is very low because it is the product of the probabilities for each individual region. For example, if the chance that two people have the same DNA sequence at any one tested region is 0.2 and 10 regions are tested, the chance that they have the same sequence at all 10 regions is 0.2^{10}, which equals approximately one in 10 million.

Chapter 4

1. Researchers obviously didn't design experimental tests of the teratogenicity of retinoic acid with human subjects. However, an acne medicine called Accutane, which contains a derivative of retinoic acid as an active ingredient, was used by some pregnant women despite warnings to the contrary. About half of one set of women who were exposed to Accutane during pregnancy either miscarried or had babies with various birth defects.

2. By contrast, imagine a more general hypothesis that some unspecified chemical in the water causes deformities. Perhaps this could be tested by finding a set of ponds where a high percentage of frogs are deformed and another set where most frogs are normal, collecting water from these ponds, and comparing the chemical constituents of these samples. But natural bodies of water differ chemically in hundreds if not thousands of ways, so it would be very difficult to pinpoint the precise chemical that might be responsible for deformities. The problem is compounded by the fact that chemical reactions take place continuously in water so the hypothetical compound that was present in the right concentration to cause abnormal limb development at the time of metamorphosis from tadpoles to frogs might have been converted into something else by the time water samples were taken. This approach is searching for a needle in a haystack.

3. There are a great variety of methods for estimating the sizes of natural populations. Some of these are fairly straightforward, such as counting the number of males displaying at a pond during the breeding season and using this as an index of abundance; others involve sophisticated mathematical analysis of resightings or recaptures of marked animals. Regardless of whether estimates come from a simple or complicated method, the key requirement is that methods for each population must be consistent over time. It's not necessary for the validity of the analysis by Houlahan et al. (2000) that all researchers used the same method, only that they didn't change how they estimated the size of a particular population in midstream.

4. Conceivably the hypothesis that the pathogenic fungus was transferred from hatchery-reared fish to frogs and toads could be tested by selecting several pristine lakes without such fish, introducing fish infected with the fungus to a random subset of these lakes, and introducing uninfected fish to the remaining lakes as controls. Then the egg survival and population dynamics of amphibians that are breeding in these lakes could be followed for several years. But the ethics of doing this experiment are questionable, considering the threatened status of some of the amphibian species.

References

Ammon, H. P. T., P. R. Bieck, D. Mandalaz, and E. J. Verspohl. 1983. Adaptation of blood pressure to continuous heavy coffee drinking in young volunteers: A double-blind crossover study. *British Journal of Clinical Pharmacology* 15:701–706.

Andrieu, N., and A. M. Goldstein. 1998. Epidemiologic and genetic approaches in the study of gene-environment interaction: An overview of available methods. *Epidemiologic Reviews* 20:137–147.

Andronova, N. G., and M. E. Schlesinger. 2001. Objective estimation of the probability density function for climate sensitivity. *Journal of Geophysical Research-Atmospheres* 106:22605–22611.

Angell, M. 1996. *Science on trial: The clash of medical evidence and the law in the breast implant case.* Norton, New York.

Ankley, G. T., J. E. Tietge, D. L. DeFoe, K. M. Jensen, G. W. Holcombe, E. J. Durham, and S. A. Diamond. 1998. Effects of ultraviolet light and methoprene on survival and development of *Rana pipiens. Environmental Toxicology and Chemistry* 17:2530–2542.

Annas, G. J. 1994. Scientific evidence in the courtroom—the death of the Frye Rule. *New England Journal of Medicine* 330:1018–1021.

Antithrombotic Trialists' Collaboration. 2002. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *British Medical Journal* 324:71–86.

Armstrong, D. P. 1991. Levels of cause and effect as organizing principles for research in animal behaviour. *Canadian Journal of Zoology* 69:823–829.

Audera, C., R. V. Patulny, B. H. Sander, and R. M. Douglas. 2001. Mega-doses of vitamin C in treatment of the common cold: A randomised controlled trial. *Medical Journal of Australia* 175:359–362.

Austad, S. N. 1993. Retarded senescence in an insular population of Virginia opossums (*Didelphis virginiana*). *Journal of Zoology, London* 229:695–708.

———. 1997. *Why we age: What science is discovering about the body's journey through life.* Wiley, New York.

Bak, A. A. A., and D. E. Grobbee. 1989. Abstinence from coffee leads to a fall in blood pressure. *Journal of Hypertension* 7:S260–S261.

Balda, R. P., and A. C. Kamil. 1989. A comparative study of cache recovery by three corvid species. *Animal Behaviour* 38:486–495.

———. 1992. Long-term spatial memory in Clark's nutcracker, *Nucifraga columbiana. Animal Behaviour* 44:761–769.

Bartlett, R. H., D. W. Roloff, J. R. Custer, J. G. Younger, and R. B. Hirschl. 2000. Extracorporeal life support: The University of Michigan experience. *JAMA* 283: 904–908.

Beckman, K. B., and B. N. Ames. 1998. The free radical theory of aging matures. *Physiological Reviews* 78:547–581.

Begg, C. B. 2001. The search for cancer risk factors: When can we stop looking? *American Journal of Public Health* 91:360–364.

Blaustein, A. R. 1994. Amphibians in a bad light. *Natural History* 103(10):32–39.

Blaustein, A. R., P. D. Hoffman, D. G. Hokit, J. M. Kiesecker, S. C. Walls, and J. B. Hays. 1994a. UV repair and resistance to solar UV-B in amphibian eggs: A link to population declines? *Proceedings of the National Academy of Sciences USA* 91: 1791–1795.

Blaustein, A. R., D. G. Hokit, R. K. O'Hara, and R. A. Holt. 1994b. Pathogenic fungus contributes to amphibian losses in the Pacific Northwest. *Biological Conservation* 67:251–254.

Blaustein, A. R., and P. T. J. Johnson. 2003. The complexity of deformed amphibians. *Frontiers in Ecology and the Environment* 1:87–94.

Blaustein, A. R., and J. M. Kiesecker. 2002. Complexity in conservation: Lessons from the global decline of amphibian populations. *Ecology Letters* 5:597–608.

Blaustein, A. R., J. M. Kiesecker, D. P. Chivers, D. G. Hokit, A. Marco, L. K. Belden, and A. Hatch. 1998. Effects of ultraviolet radiation on amphibians: Field experiments. *American Zoologist* 38:799–812.

Blaustein, A. R., J. M. Kiesecker, D. G. Hokit, and S. C. Walls. 1995. Amphibian declines and UV radiation. *Bioscience* 45:514–515.

Blaustein, A. R., and D. B. Wake. 1995. The puzzle of declining amphibian populations. *Scientific American* 272(4):52–57.

Bouma, M. J., C. Dye, and H. J. Van Der Kaay. 1996. Falciparum malaria and climate change in the Northwest Frontier Province of Pakistan. *American Journal of Tropical Medicine and Hygiene* 55:131–137.

Braun, B. L., J. B. Fowles, L. Solberg, E. Kind, M. Healey, and R. Anderson. 2000. Patient beliefs about the characteristics, causes, and care of the common cold: An update. *Journal of Family Practice* 49:153–156.

Brisbin, I. L., Jr., and S. N. Austad. 1991. Testing the individual odour theory of canine olfaction. *Animal Behaviour* 42:63–69.

Brisbin, I. L., Jr., S. Austad, and S. K. Jacobson. 2000. Canine detectives: The nose knows — or does it? *Science* 290:1093.

Bruton, S., and F. Fong. 2000. *Science content standards for California public schools: Kindergarten through Grade 12.* California Department of Education, Sacramento.

Burr, M. L., J. E. J. Gallacher, B. K. Butland, C. H. Bolton, and L. G. Downs. 1989. Coffee, blood pressure and plasma lipids: A randomized controlled trial. *European Journal of Clinical Nutrition* 43:477–483.

Carter, K. C. 1985. Koch's postulates in relation to the work of Jacob Henle and Edwin Klebs. *Medical History* 29:353–374.

Chalmers, T. C. 1975. Effects of ascorbic acid on the common cold: An evaluation of the evidence. *American Journal of Medicine* 58:532–536.

———. 1996. Dissent to the preceding article by H. Hemilä. *Journal of Clinical Epidemiology* 49:1085.

Chamberlin, T. C. 1890. The method of multiple working hypotheses. *Science* (old series) 15:92–96 (reprinted in *Science* 148:754–759 [1965]).

Chinni, D. 2001. Arsenic flap and 'sound science.' *Christian Science Monitor,* 14 June 2001.

Christie, M. 2000. *The ozone layer: A philosophy of science perspective.* Cambridge University Press, Cambridge.

Clausen, J., D. D. Keck, and W. M. Hiesey. 1948. *Experimental studies on the nature of species. III. Environmental responses of climatic races of Achillea.* Carnegie Institution of Washington Publication No. 581, Washington, D.C.

Cohen, J. 2003. Vaccine results lose significance under scrutiny. *Science* 299:1495.

Colditz, G., W. Dejong, D. Hunter, D. Trichopoulos, and W. Willett. 1996. Harvard report on cancer prevention, volume 1: Causes of human cancer. *Cancer Causes and Control* 7:S3–S59.

Colditz, G. A., W. Dejong, K. Emmons, D. J. Hunter, N. Mueller, and G. Sorensen. 1997. Harvard report on cancer prevention, volume 2: Prevention of human cancer. *Cancer Causes and Control* 8:S1–S50.

Congdon, J. D., R. D. Nagle, O. M. Kinney, and R. C. van Loben Sels. 2001. Hypotheses of aging in a long-lived vertebrate, Blanding's turtle (*Emydoidea blandingii*). *Experimental Gerontology* 36:813–827.

Culotta, E. 1991. Science's 20 greatest hits take their lumps. *Science* 251:1308–1309.

De Angelis, H., and P. Skvarca. 2003. Glacier surge after ice shelf collapse. *Science* 299:1560–1562.

Douglas, R. M., E. B. Chalker, and B. Treacy. 2001. *Vitamin C for preventing and treating the common cold (Cochrane Review).* The Cochrane Library, Issue 4, Update Software, Oxford.

Draper, G. J., B. M. Sanders, P. A. Brownbill, and M. M. Hawkins. 1992. Patterns of risk of hereditary retinoblastoma and applications to genetic counselling. *British Journal of Cancer* 66:211–219.

Dupont, W. D., and W. D. Plummer, Jr. 1996. Understanding the relationship between relative and absolute risk. *Cancer* 77:2193–2199.

Eggertsen, R., A. Andreasson, T. Hedner, B. E. Karlberg, and L. Hansson. 1993. Effect of coffee on ambulatory blood pressure in patients with treated hypertension. *Journal of Internal Medicine* 233:351–355.

Eisenberg, J. F., and D. E. Wilson. 1978. Relative brain size and feeding strategies in the Chiroptera. *Evolution* 32:740–751.

Ekbom, A. 1999. Commentary [Review: Substantial coffee consumption was associated with a lower risk of colorectal cancer in the general population]. *Gut* 44:597.

Enserink, M. 2002. The vanishing promises of hormone replacement. *Science* 297:325–326.

Epstein, P. R. 2000. Is global warming harmful to health? *Scientific American* 283(2): 50–57.

———. 2001. Climate change and emerging infectious diseases. *Microbes and Infection* 3:747–754.

Evans, P., and B. Halliwell. 2001. Micronutrients: Oxidant/antioxidant status. *British Journal of Nutrition* 85:S67–S74.

Ewing, R. C., and A. Macfarlane. 2002. Yucca Mountain. *Science* 296:659–660.

Faddy, M. J., R. G. Gosden, A. Gougeon, S. J. Richardson, and J. F. Nelson. 1992. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. *Human Reproduction* 7:1342–1346.

Finch, C. E., and T. B. L. Kirkwood. 2000. *Chance, development, and aging.* Oxford University Press, New York.

Finch, C. E., M. C. Pike, and M. Witten. 1990. Slow mortality rate accelerations during aging in some animals approximate that of humans. *Science* 249:902–905.

Food and Drug Administration, U.S. Department of Health and Human Services. 2002. *Transcript of meeting of the Antiviral Drugs Advisory Committee on 19 March 2002.* http://www.fda.gov/ohrms/dockets/ac/02/transcripts/3847t1.pdf (accessed April 1, 2003).

Food and Nutrition Board, Institute of Medicine, U.S. National Academy of Sciences (Panel on Dietary Antioxidants and Related Compounds, Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes and Standing Committee on the Scientific Evaluation of Dietary Reference Intakes). 2000. *Dietary reference intakes for vitamin C, vitamin E, selenium, and carotenoids.* National Academy Press, Washington, D. C.

Gannett, L. 1999. What's in a cause?: The pragmatic dimensions of genetic explanations. *Biology & Philosophy* 14:349–374.

Gaulin, S. J. C., and R. W. FitzGerald. 1986. Sex differences in spatial ability: An evolutionary hypothesis and test. *The American Naturalist* 127:74–88.

Giere, R. N. 1997. *Understanding scientific reasoning*, 4th ed. Harcourt Brace College Publishers, Fort Worth, Tex.

———. 1999. *Science without laws.* University of Chicago Press, Chicago.

Giovannucci, E. 1998. Meta-analysis of coffee consumption and risk of colorectal cancer. *American Journal of Epidemiology* 147:1043–1052.

———. 1999. Review: Substantial coffee consumption was associated with a lower risk of colorectal cancer in the general population. *Gut* 44:597.

Glass, G. V. 1977. Integrating findings: The meta-analysis of research. *Review of Research in Education* 5:351–379.

Gomulkiewicz, R., and N. A. Slade. 1997. Legal standards and the significance of DNA evidence. *Human Biology* 69:675–688.

Gould, J. L. 1982. *Ethology: The mechanisms and evolution of behavior.* Norton, New York.

Gould, S. J. 1997. Nonoverlapping Magisteria. *Natural History* 106(2):16–22, 60–62.

Graudal, N. A., A. M. Galloe, and P. Garred. 1998. Effects of sodium restriction on blood pressure, renin, aldosterone, catecholamines, cholesterols, and triglyceride: A meta-analysis. *JAMA* 279:1383–1391.

Griffin, D. R. 1986. *Listening in the dark: The acoustic orientation of bats and men.* Cornell University Press, Ithaca, N.Y.

Grimes, D. A., and K. F. Schulz. 2002. Uses and abuses of screening tests. *The Lancet* 359:881–884.

Hardin, G. 1968. The tragedy of the commons. *Science* 162:1243–1248.

———. 1976. Vulnerability—the strength of science. *American Biology Teacher* 38: 465, 483.

Harman, D. 1956. Aging: A theory based on free radical and radiation chemistry. *Journal of Gerontology* 11:298–300.

Hartman, T. J., J. A. Tangrea, P. Pietinen, N. Malila, M. Virtanen, P. R. Taylor, and

D. Albanes. 1998. Tea and coffee consumption and risk of colon and rectal cancer in middle-aged Finnish men. *Nutrition and Cancer* 31:41–48.

Hayes, T., K. Haston, M. Tsui, A. Hoang, C. Haeffele, and A. Vonk. 2002. Feminization of male frogs in the wild. *Nature* 419:895–896.

Hayflick, L., and P. S. Moorhead. 1961. The serial cultivation of human diploid cell strains. *Experimental Cell Research* 25:585–621.

Hazen, R. M., and J. Trefil. 1991. *Science matters: Achieving scientific literacy.* Doubleday, New York.

Hemilä, H. 1996. Vitamin C, the placebo effect, and the common cold: A case study of how preconceptions influence the analysis of results. *Journal of Clinical Epidemiology* 49:1079–1084.

Hepper, P. G. 1988. The discrimination of human odour by the dog. *Perception* 17: 549–554.

Hilborn, R., and S. C. Stearns. 1982. On inference in ecology and evolutionary biology: The problem of multiple causes. *Acta Biotheoretica* 31:145–164.

Hoffrage, U., S. Lindsey, R. Hertwig, and G. Gigerenzer. 2000. Communicating statistical information. *Science* 290:2261–2262.

Holekamp, K. E., and P. W. Sherman. 1989. Why male ground squirrels disperse. *American Scientist* 77:232–239.

Holmes, D. J., and S. N. Austad. 1995. Birds as animal models for the comparative biology of aging: A prospectus. *Journal of Gerontology* 50A:B59–B66.

Hoover, R. N. 2000. Cancer—nature, nurture, or both. *New England Journal of Medicine* 343:135–136.

Houlahan, J. E., C. S. Findlay, B. R. Schmidt, A. H. Meyer, and S. L. Kuzmin. 2000. Quantitative evidence for global amphibian population declines. *Nature* 404: 752–755.

Hróbjartsson, A., and P. C. Gøtzsche. 2001. Is the placebo powerless? *New England Journal of Medicine* 344:1594–1602.

Hutchins, H. E., and R. M. Lanner. 1982. The central role of Clark's nutcracker in the dispersal and establishment of whitebark pine. *Oecologia* 55:192–201.

IPCC. 2001. *Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge and New York.

Jacobs, L. F., S. J. C. Gaulin, D. F. Sherry, and G. E. Hoffman. 1990. Evolution of spatial cognition: Sex-specific patterns of spatial behavior predict hippocampal size. *Proceedings of the National Academy of Sciences, USA* 87:6349–6352.

Jacobs, L. F., and W. D. Spencer. 1994. Natural space-use patterns and hippocampal size in kangaroo rats. *Brain, Behavior and Evolution* 44:125–132.

Jasanoff, S. 1995. *Science at the bar: Law, science, and technology in America.* Harvard University Press, Cambridge, Mass.

Jee, S. H., J. He, P. K. Whelton, I. Suh, and M. J. Klag. 1999. The effect of chronic coffee drinking on blood pressure: A meta-analysis of controlled clinical trials. *Hypertension* 33:647–652.

Johansen, B. E. 2002. *The global warming desk reference.* Greenwood Press, Westport, Conn.

Johnson, P. T. J., K. B. Lunde, E. G. Ritchie, and A. E. Launer. 1999. The effect of trematode infection on amphibian limb development and survivorship. *Science* 284:802–804.

Johnson, P. T. J., K. B. Lunde, E. M. Thurman, E. G. Ritchie, S. N. Wray, D. R. Sutherland, J. M. Kapfer, T. J. Frest, J. Bowerman, and A. R. Blaustein. 2002.

Parasite (*Ribeiroia ondatrae*) infection linked to amphibian malformations in the western United States. *Ecological Monographs* 72:151–168.

Kaiser, J. 2001. Second look at arsenic finds higher risk. *Science* 293:2189.

Karlowski, T. R., T. C. Chalmers, L. D. Frenkel, A. Z. Kapikian, T. L. Lewis, and J. M. Lynch. 1975. Ascorbic acid for the common cold: A prophylactic and therapeutic trial. *JAMA* 231:1038–1042.

Kemper, A. R., M. M. Davis, and G. L. Freed. 2002. Expected adverse events in a mass smallpox vaccination campaign. *Effective Clinical Practice* 5:84–90.

Kiesecker, J. M. 2002. Synergism between trematode infection and pesticide exposure: A link to amphibian limb deformities in nature? *Proceedings of the National Academy of Sciences USA* 99:9900–9904.

Kiesecker, J. M., and A. R. Blaustein. 1995. Synergism between UV-B radiation and a pathogen magnifies amphibian embryo mortality in nature. *Proceedings of the National Academy of Sciences USA* 92:11049–11052.

Kiesecker, J. M., A. R. Blaustein, and L. K. Belden. 2001. Complex causes of amphibian population declines. *Nature* 410:681–684.

Klag, M. J., N.-Y. Wang, L. A. Meoni, F. L. Brancati, L. A. Cooper, K.-Y. Liang, J. H. Young, and D. E. Ford. 2002. Coffee intake and risk of hypertension: The Johns Hopkins Precursors Study. *Archives of Internal Medicine* 162:657–662.

Krebs, J. R. 1990. Food-storing birds: Adaptive specialization in brain and behaviour? *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences* 329:153–160.

Krebs, J. R., D. F. Sherry, S. D. Healy, V. H. Perry, and A. L. Vaccarino. 1989. Hippocampal specialization of food-storing birds. *Proceedings of the National Academy of Sciences USA* 86:1388–1392.

Lanner, R. M., and K. F. Connor. 2001. Does bristlecone pine senesce? *Experimental Gerontology* 36:675–685.

Lee, D. W., L. E. Miyasato, and N. S. Clayton. 1998. Neurobiological bases of spatial learning in the natural environment: Neurogenesis and growth in the avian and mammalian hippocampus. *NeuroReport* 9:R15–R27.

LeLorier, J., G. Grégoire, A. Benhaddad, J. Lapierre, and F. Derderian. 1997. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine* 337:536–542.

Licht, L. E. 1995. Disappearing amphibians? *Bioscience* 45:307.

———. 1996. Amphibian decline still a puzzle. *Bioscience* 46:172–173.

Lichtenstein, P., N. V. Holm, P. K. Verkasalo, A. Iliadou, J. Kaprio, M. Koskenvuo, E. Pukkala, A. Skytthe, and K. Hemminki. 2000. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine* 343:78–85.

Light, R. J., and D. B. Pillemer. 1984. *Summing up: The science of reviewing research.* Harvard University Press, Cambridge, Mass.

Linet, M. S. 2000. Evolution of cancer epidemiology. *Epidemiologic Reviews* 22:35–56.

Lorentz, C. P., E. D. Wieben, A. Tefferi, D. A. H. Whiteman, and G. W. Dewald. 2002. Primer on medical genomics. Part I: History of genetics and sequencing of the human genome. *Mayo Clinic Proceedings* 77:773–782.

Macdonald, G. 1961. Epidemiologic models in studies of vector-borne diseases. *Public Health Reports* 76:753–764.

MacDonald, T. M., K. Sharpe, G. Fowler, D. Lyons, S. Freestone, H. G. Lovell, J. Webster, and J. C. Petrie. 1991. Caffeine restriction: Effect on mild hypertension. *British Medical Journal* 303:1235–1238.

Macklin, R. 1999. Ethics, epidemiology, and law: The case of silicone breast implants. *American Journal of Public Health* 89:487–489.

Maguire, E. A., D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. J. Frackowiak, and C. D. Frith. 2000. Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences USA* 97: 4398–4403.

Mahlman, J. D. 1998. Science and nonscience concerning human-caused climate warming. *Annual Review of Energy and the Environment* 23:83–105.

Martens, P. 1999. How will climate change affect human health? *American Scientist* 87:534–541.

Martens, P., R. S. Kovats, S. Nijhof, P. de Vries, M. T. J. Livermore, D. J. Bradley, J. Cox, and A. J. McMichael. 1999. Climate change and future populations at risk of malaria. *Global Environmental Change* 9:S89–S107.

Marx, J. 2002. Chromosome end game draws a crowd. *Science* 295:2348–2351.

May, R. M. 1983. Parasitic infections as regulators of animal populations. *American Scientist* 71:36–45.

Mayr, E. 1961. Cause and effect in biology. *Science* 134:1501–1506.

McCauliff, C. M. A. 1982. Burdens of proof: Degrees of belief, quanta of evidence, or constitutional guarantees? *Vanderbilt Law Review* 35:1293–1335.

McGue, M., J. W. Vaupel, N. Holm, and B. Harvald. 1993. Longevity is moderately heritable in a sample of Danish twins born 1870–1880. *Journal of Gerontology* 48:B237–B244.

Medawar, P. B. 1952. An unsolved problem of biology. H. K. Lewis, London.

Medvedev, Z. A. 1990. An attempt at a rational classification of theories of ageing. *Biological Reviews* 65:375–398.

Meydani, M. 2001. Antioxidants and cognitive function. *Nutrition Reviews* 59:S75–S82.

Murphree, A. L. 1997. Retinoblastoma. In D. L. Rimoin, J. M. Connor, and R. E. Pyeritz, eds., *Emery and Rimoin's principles and practice of medical genetics*, pp. 2585–2609. Churchill Livingstone, New York.

National Science Board. 1998. *Science and engineering indicators—1998*. National Science Foundation, Arlington, Virginia.

Newman, B., R. C. Millikan, and M.-C. King. 1997. Genetic epidemiology of breast and ovarian cancers. *Epidemiologic Reviews* 19:69–79.

Nottebohm, F. 1981. A brain for all seasons: Cyclical anatomical changes in song control nuclei of the canary brain. *Science* 214:1368–1370.

O'Brien, S. J., and M. Dean. 1997. In search of AIDS-resistance genes. *Scientific American* 277(3):44–51.

Park, R. L. 2000. *Voodoo science: The road from foolishness to fraud*. Oxford University Press, New York.

Pauling, L. 1970. *Vitamin C and the common cold*. W. H. Freeman, San Francisco.

———. 1971. The significance of the evidence about ascorbic acid and the common cold. *Proceedings of the National Academy of Sciences USA* 68:2678–2681.

Pennisi, E. 2001. The human genome. *Science* 291:1177–1180.

Perls, T. T., L. Alpert, and R. C. Fretts. 1997. Middle-aged mothers live longer. *Nature* 389:133.

Perls, T. T., E. Bubrick, C. G. Wager, J. Vijg, and L. Kruglyak. 1998. Siblings of centenarians live longer. *The Lancet* 351:1560.

Perrig, W. J., P. Perrig, and H. B. Stähelin. 1997. The relation between antioxidants and memory performance in the old and very old. *Journal of the American Geriatrics Society* 45:718–724.

Peto, R. 2000. Cancer, genes, and the environment. *New England Journal of Medicine* 343:1495.

Piedbois, P., and M. Buyse. 2000. Recent meta-analyses in colorectal cancer. *Current Opinion in Oncology* 12:362–367.

Platt, J. R. 1964. Strong inference. *Science* 146:347–353.

Pool, R. 1991. Science literacy: The enemy is us. *Science* 251:266–267.

Popper, K. R. 1965. *Conjectures and refutations: The growth of scientific knowledge.* Basic Books, New York.

Pravosudov, V. V., and N. S. Clayton. 2002. A test of the adaptive specialization hypothesis: Population differences in caching, memory, and the hippocampus in black-capped chickadees (*Poecile atricapilla*). *Behavioral Neuroscience* 116:515–522.

Puca, A. A., M. J. Daly, S. J. Brewster, T. C. Matise, J. Barrett, M. Shea-Drinkwater, S. Kang, E. Joyce, J. Nicoli, E. Benson, L. M. Kunkel, and T. Perls. 2001. A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proceedings of the National Academy of Sciences USA* 98:10505–10508.

Reboreda, J. C., N. S. Clayton, and A. Kacelnik. 1996. Species and sex differences in hippocampus size in parasitic and non-parasitic cowbirds. *Neuroreport* 7:505–508.

Reeburgh, W. S. 1997. *Figures summarizing the global cycles of biologically active elements.* http://ess1.ps.uci.edu/~reeburgh/figures.html (accessed April 3, 2003).

Rensberger, B. 2000. The nature of evidence. *Science* 289:61.

Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* 43:223–225.

Ricklefs, R. E. 1998. Evolutionary theories of aging: Confirmation of a fundamental prediction, with implications for the genetic basis and evolution of life span. *The American Naturalist* 152:24–44.

Ricklefs, R. E., and C. E. Finch. 1995. *Aging: A natural history.* W. H. Freeman, New York.

Risch, N. 2001. The genetic epidemiology of cancer: Interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiology, Biomarkers & Prevention* 10:733–741.

Robinson, D. 2002. *Cancer clusters: Findings vs feelings.* Medscape General Medicine 4(4). http://www.medscape.com/viewarticle/442554 (accessed November 23, 2002).

Roeder, K. 1994. DNA fingerprinting: A review of the controversy. *Statistical Science* 9:222–247.

Rogers, D. J., and S. E. Randolph. 2000. The global spread of malaria in a future, warmer world. *Science* 289:1763–1766.

Rose, G. 1981. Strategies of prevention: Lessons from cardiovascular disease. *British Medical Journal* 282:1847–1851.

———. 1992. *The strategy of preventive medicine.* Oxford University Press, Oxford.

Rose, M. R. 1984. Laboratory evolution of postponed senescence in *Drosophila melanogaster. Evolution* 38:1004–1010.

Rosmarin, P. C., W. B. Applegate, and G. W. Somes. 1990. Coffee consumption and blood pressure: A randomized, crossover clinical trial. *Journal of General Internal Medicine* 5:211–213.

Royall, R. M. 1991. Ethics and statistics in randomized clinical trials (with discussion). *Statistical Science* 6:52–88.

Sano, M., C. Ernesto, R. G. Thomas, M. R. Klauber, K. Schafer, M. Grundman, P. Woodbury, J. Growdon, C. W. Cotman, E. Pfeiffer, L. S. Schneider, and L. J. Thal. 1997. A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. *New England Journal of Medicine* 336:1216–1222.

Sarkar, S. 1996. Ecological theory and anuran declines. *Bioscience* 46:199–207.

Schoon, G. A. A. 1996. Scent identification lineups by dogs (*Canis familiaris*): Experimental design and forensic application. *Applied Animal Behaviour Science* 49: 257–267.

———. 1997. Scent identification by dogs (*Canis familiaris*): A new experimental design. *Behaviour* 134:531–550.

———. 1998. A first assessment of the reliability of an improved scent identification line-up. *Journal of Forensic Science* 43:70–75.

Senior, K. 2002. FDA panel rejects common cold treatment. *The Lancet Infectious Diseases* 2:264.

Sessions, S. K., and S. B. Ruth. 1990. Explanation for naturally occurring supernumerary limbs in amphibians. *Journal of Experimental Zoology* 254:38–47.

Settle, R. H., B. A. Sommerville, J. McCormick, and D. M. Broom. 1994. Human scent matching using specially trained dogs. *Animal Behaviour* 48:1443–1448.

Shanley, D. P., and T. B. L. Kirkwood. 2001. Evolution of the human menopause. *BioEssays* 23:282–287.

Sharpe, D. 1997. Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review* 17:881–901.

Sherman, P. W. 1988. The levels of analysis. *Animal Behaviour* 36:616–619.

———. 1998. The evolution of menopause. *Nature* 392:759–760.

Sherman, P. W., and S. M. Flaxman. 2001. Protecting ourselves from food. *American Scientist* 89:142–151.

———. 2002. Nausea and vomiting of pregnancy in an evolutionary perspective. *American Journal of Obstetrics and Gynecology* 186:S190–S197.

Sherry, D. F., A. L. Vaccarino, K. Buckenham, and R. S. Herz. 1989. The hippocampal complex of food-storing birds. *Brain, Behavior and Evolution* 34:308–317.

Shettleworth, S. J. 1990. Spatial memory in food-storing birds. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences* 329:143–151.

Sinclair, A. J., A. J. Bayer, J. Johnston, C. Warner, and S. R. Maxwell. 1998. Altered plasma antioxidant status in subjects with Alzheimer's disease and vascular dementia. *International Journal of Geriatric Psychiatry* 13:840–845.

Souder, W. 2000. *A plague of frogs: The horrifying true story.* Hyperion, New York.

Steel, C. M. 1997. Cancer of the breast and female reproductive tract. In D. L. Rimoin, J. M. Connor, and R. E. Pyeritz, eds., *Emery and Rimoin's principles and practice of medical genetics*, 3rd ed., pp. 1501–1523. Churchill Livingstone, New York.

Stein, Z. A. 1999. Silicone breast implants: Epidemiological evidence of sequelae. *American Journal of Public Health* 89:484–487.

Stevens, T. A., and J. R. Krebs. 1986. Retrieval of stored seeds by marsh tits *Parus palustris* in the field. *Ibis* 128:513–525.

Superko, H. R., W. Bortz, Jr., P. T. Williams, J. J. Albers, and P. D. Wood. 1991. Caffeinated and decaffeinated coffee effects on plasma lipoprotein cholesterol, apolipoproteins, and lipase activity: A controlled, randomized trial. *American Journal of Clinical Nutrition* 54:599–605.

Superko, H. R., J. Myll, C. DiRicco, P. T. Williams, W. M. Bortz, and P. D. Wood. 1994. Effects of cessation of caffeinated-coffee consumption on ambulatory and resting blood pressure in men. *American Journal of Cardiology* 73:780–784.

Sutera, R. 2000. The origin of whales and the power of independent evidence. *Reports of the National Center for Science Education* 20(5):33–41.

Swift, M., D. Morrell, R. B. Massey, and C. L. Chase. 1991. Incidence of cancer in 161 families affected by ataxia telangiectasia. *New England Journal of Medicine* 325:1831–1836.

Taslitz, A. E. 1990. Does the cold nose know? The unscientific myth of the dog scent lineup. *Hastings Law Journal* 42:15–134.

Taubes, G. 1998. The (political) science of salt. *Science* 281:898–907.

Tavani, A., and C. La Vecchia. 2000. Coffee and cancer: A review of epidemiological studies, 1990–1999. *European Journal of Cancer Prevention* 9:241–256.

Taylor, K. 1999. Rapid climate change. *American Scientist* 87:320–327.

Terry, P., L. Bergkvist, L. Holmberg, and A. Wolk. 2001. Coffee consumption and risk of colorectal cancer in a population based prospective cohort of Swedish women. *Gut* 49:87–90.

Thewissen, J. G. M. 1998. *The emergence of whales: Evolutionary patterns in the origin of Cetacea.* Plenum, New York.

Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift fur Tierpsychologie* 20:410–433.

Trichopoulos, D., F. P. Li, and D. J. Hunter. 1996. What causes cancer? *Scientific American* 275(3):80–87.

Tudge, C. 2002. Why science should warm our hearts. *Reports of the National Center for Science Education* 22(1–2):41–44.

Tyner, S. D., S. Venkatachalam, J. Choi, S. Jones, N. Ghebranious, H. Igelmann, X. Lu, G. Soron, B. Cooper, C. Brayton, S. H. Park, T. Thompson, G. Karsenty, A. Bradley, and L. A. Donehower. 2002. p53 mutant mice that display early ageing-associated phenotypes. *Nature* 415:45–53.

United States Census Bureau. 2002. Vital statistics No. 98. Expectation of life and expected deaths by race, sex, and age: 1998. In 2001 *Statistical Abstract of the United States, Section 2*, p. 74. http://www.census.gov/prod/2002pubs/01statab/vitstat.pdf (accessed April 3, 2003).

van Dusseldorp, M., P. Smits, J. W. M. Lenders, L. Temme, T. Thien, and M. B. Katan. 1992. Effects of coffee on cardiovascular responses to stress: A 14-week controlled trial. *Psychosomatic Medicine* 54:344–353.

van Dusseldorp, M., P. Smits, T. Thien, and M. B. Katan. 1989. Effect of decaffeinated coffee versus regular coffee on blood pressure: A 12-week, double-blind trial. *Hypertension* 14:563–569.

Vander Wall, S. B. 1982. An experimental analysis of cache recovery in Clark's nutcracker. *Animal Behaviour* 30:84–94.

———. 1990. *Food hoarding in animals.* University of Chicago Press, Chicago.

———. 1998. Foraging success of granivorous rodents: Effects of variation in seed and soil water on olfaction. *Ecology* 79:233–241.

———. 2000. The influence of environmental conditions on cache recovery and cache pilferage by yellow pine chipmunks (*Tamias amoenus*) and deer mice (*Peromyscus maniculatus*). *Behavioral Ecology* 11:544–549.

Vander Wall, S. B., and R. P. Balda. 1977. Coadaptations of the Clark's nutcracker and the piñon pine for efficient seed harvest and dispersal. *Ecological Monographs* 47:89–111.

Varmus, H., and R. A. Weinberg. 1993. *Genes and the biology of cancer.* Freeman, New York.

Vogel, F., and A. G. Motulsky. 1997. *Human genetics: Problems and approaches*, 3rd ed. Springer-Verlag, Berlin.

Ware, J. H. 1989. Investigating therapies of potentially great benefit: ECMO (with discussion). *Statistical Science* 4:298–340.

Weinberg, R. A. 1996. How cancer arises. *Scientific American* 275(3):62–71.

Wells, D. L., and P. G. Hepper. 2000. The discrimination of dog odours by humans. *Perception* 29:111–115.

Wells, G. L., R. S. Malpass, R. C. L. Lindsay, R. P. Fisher, J. W. Turtle, and S. M. Fulero. 2000. From the lab to the police station—a successful application of eyewitness research. *American Psychologist* 55:581–598.

Willett, W. C. 2002. Balancing life-style and genomics research for disease prevention. *Science* 296:695–698.

Wilson, E. O. 2002. *The future of life.* Knopf, New York.

Wong, K. 2002. The mammals that conquered the seas. *Scientific American* 286(5): 70–79.

Yang, Q., and M. J. Khoury. 1997. Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. *Epidemiologic Reviews* 19:33–43.

APPENDIX D

Anim. Behav., 1991, **42**, 63–69

Testing the individual odour theory of canine olfaction

I. LEHR BRISBIN, JR* & STEVEN N. AUSTAD†‡

**Savannah River Ecology Laboratory, P.O. Drawer E, Aiken, South Carolina 29801, U.S.A.*

†Department of Organismic and Evolutionary Biology, The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, U.S.A.

(Received 6 August 1990; initial acceptance 28 September 1990; final acceptance 29 November 1990; MS. number: A5844)

Abstract. Experiments were performed with three dogs, *Canis familiaris*, trained in human scent discrimination (American Kennel Club utility obedience test), to evaluate whether the dogs could distinguish the scent of their handler from the scent of other humans, irrespective of the body part from which the scent had been collected. The dogs were successful at distinguishing scent obtained from the hand of their handler from that from the hands of strangers, but could not similarly distinguish their handler's scent when it was obtained from the crook of his arm. These results suggest either that there is no such thing as an individual human odour or that dogs trained with standard methods do not spontaneously identify individual odour components of scents taken from different parts of the body. The results also call into question the practice of using dogs to identify individuals from scented objects in law enforcement, unless the dogs used can be shown to be capable of performing discriminations of the type unsuccessfully attempted by the animals in the present study.

The ability to identify the scent of particular individuals is part of the folk wisdom associated with canine olfactory abilities. In keeping with this belief, domestic dogs, *Canis familiaris*, are often used in law enforcement to attempt to match pieces of physical evidence with presumed criminal perpetrators. Implicit in this practice is the assumption that humans have a single 'individual scent', which can be identified regardless of the body part from which the odour comes. For instance, an individual's shirt presumably would be recognized as having the same identifying scent as his handkerchief, socks or a utensil touched while eating. Our aim was to provide a controlled experimental evaluation of this hypothesis about canine olfactory ability and individual human odours. At the same time we hoped that the study might yield insight into the information inherent in mammalian scent.

In the United States, the American Kennel Club sanctions competitive dog obedience competition in which licensed judges evaluate the abilities of trained dogs to perform certain tasks (American Kennel Club 1979). In one of these tasks the dog is required to select from 10 similar 'scent articles' (five of which are made of leather and five of metal), the one leather and one metal article scented from the palm of its handler's hand, the other eight articles being similarly scented from the palm of the judge's hand. In this test, the articles are placed in a group on the ground at least 3 m from the dog and handler. The dog is then required to leave the handler and go to the articles where it must work independently, select the correct article, pick it up in its mouth and bring it back to the handler. Dogs that successfully perform this and several other obedience tests (unrelated to scent discrimination) on three different occasions under at least two different judges, are awarded the title 'Utility Dog' by the American Kennel Club. Such Utility Dogs, or those that have been trained to a level of proficiency sufficient to qualify in at least the scent discrimination section of these tests, thus provide excellent subjects on which to test scent discrimination abilities.

Dogs trained to compete in Utility Dog scent discrimination tests have usually had a minimum of 6 months to 1 year of training and practice in this testing format, and can be assumed to have never been exposed to scent-discrimination decisions involving any scent other than that imparted by the palm of the hand. Under American Kennel Club

---

‡To whom correspondence should be addressed.

0003–3472/91/070063+07 $03.00/0

regulations the dog can be disqualified if during the obedience competition scent is imparted to the article by any other part of the body.

Using three dogs with this experience, we have investigated whether they will identify articles scented from a different part of their handler's body without further training. If dogs with this training use a single odour to identify each individual, they should be able to distinguish their owner's scent regardless of the body part used to impart the scent.

METHODS

Experimental Subjects

Three dogs (Table I) were tested extensively between 1983 and 1987. A less complete series of additional experiments was undertaken with one of these dogs, BH, in 1989, in an attempt to clarify certain aspects of the previous results.

Training Method

The training method was loosely based on that of Bauman (1986). Each dog was initially trained to retrieve standard training dumbbells on command without respect to scent. Thereafter, the dog was asked to retrieve a dumbbell scented by the hand of the handler and placed on a pegboard near an identical 'unscented' dumbbell, which was wired to the board. The unscented dumbbell was one that had been untouched by humans for at least 24 h, and had been stored in a tray open to the air during that time. When the dogs had learnt that the unscented dumbbell could not be retrieved (because it was wired to the board) and had therefore become used to retrieving the scented one, the unscented dumbbell was unwired.

When both dumbbells were free to be retrieved, dogs retrieving the wrong (i.e. unscented) dumbbell were made to repeat the process until the right dumbbell was retrieved, at which time to dog was praised, and allowed to stop. Dogs retrieving the correct dumbbell on their first try were praised and not required to repeat the retrieval.

Gradually, more and more identical dumbbells were added to the pegboard, some unscented, some scented by the hands of other people. The new 'scent' objects were always wired down until the dogs learned to ignore them, after which they were unwired.

Experimental Protocol

Testing procedures were based on the design of the Utility Dog Scent Discrimination Test for which all of the dogs had originally been trained. The scent articles used were also the ones customarily used in Utility Dog tests. All articles were made in the shape of dumbbells consisting of cylindrical barrels measuring 10×2 cm connecting 7·5-cm^2 ends. Each article was individually numbered to assure identification. During all tests the articles were arranged so that the dog's handler could not see the numbers and was therefore unaware of which article bore his scent while the dog was making a choice.

Scented articles were presented to the dog in pairs. The dumbbells were separated from each other by about 5 cm, with their barrels parallel and in a line perpendicular to the path by which the dog approached after being sent by the handler. Articles were placed in the centre of a 360-cm^2 board, which was placed on the ground 3–4 m in front of the dog and handler.

In conducting the test, the dog was first placed in a sitting 'heel' position by the handler's left side facing away from the board on which the articles were to be placed. The experimenter (as distinct from the handler) arranged to have the appropriate scent imparted to the articles as explained below. Then, using 23-cm-long tongs, the experimenter placed the articles on the board.

Left and right positions were assigned at random. Neither the dog nor the handler could see the two articles being placed on the board. After approximately 30 s, the dog and handler performed an about-turn to face the board, with the dog returning to the sitting position by the handler's side. The dog was then sent forward to the board by a command from the handler. No additional commands or signals were given once the dog began moving towards the board. Once at the board, the dog selected one of the articles and brought it back to the handler. In addition to recording which article was selected by the dog, the experimenter recorded the sequence and number of times the dog sniffed the right and left articles. A sniff was recorded when the dog extended its nose to within 4–5 cm of the central barrel of one of the articles. In the 1989 tests using dog BH, the time elapsing between arriving at the board and making the choice (i.e. picking up the article in its mouth) was also recorded.

Scent was imparted to the articles by one of two methods. Scent designated as 'hand' was imparted by having the appropriate person hold the barrel of the article in the closed palm of the hand for 30 s.

Table 1. Dogs used in scent discrimination tests

Code	Breed	Sex	Year of birth	Titles*/training level
LR	Labrador retriever	Male	1972	CD, CDX, UD, TD
AST	American Staffordshire terrier†	Male	1975	CD, CDX/trained to level of competitive proficiency in all UD exercises‡
BH	Bloodhound	Male§	1978	CD, CDX/trained to level of competitive proficiency in all UD exercises‡

*CD: Companion Dog; CDX: Companion Dog Excellent; UD: Utility Dog; TD: Tracking Dog. Requirements to attain the CD, CDX, UD and TD titles are described in detail by American Kennel Club (1979). Although the TD title requires the dog to track human scent under field conditions, only the UD title requires scent discrimination performance as described in the text.

†Also registered as an American pit bull terrier.

‡Dog trained to a level of proficiency sufficient to allow it to compete in UD obedience competition although the title had not yet actually been obtained.

§This dog was also used in man-trailing work for purposes of law enforcement. However, no scent discrimination training had been given other than that related to the UD tests.

Alternatively, 'elbow' scent was imparted by using tongs to position the article in the crease of the antecubital space of the subject's forearm (crook of the arm). The subject then clamped the article tightly against the skin of the antecubital space by flexing the forearm and bringing the hand to the shoulder and holding this position for 30 s. Both hand and elbow scent were obtained from the same side of the subject's body. At no time were the ends of the dumbbell touched by the subjects, and dogs almost always seemed to obtain scent from the articles by sniffing directly at the middle of the barrel. For controls, 'no scent' was obtained by handling the article only with tongs, such that there was no contact with any human skin during the duration of the experiment.

Following testing, the barrels of all articles were lightly rubbed with the palm of the handler's hand in an attempt to decrease any scent differential between them. They were then placed in an open tray and exposed to air for at least 24 h before being used again in tests. General experience in training and handling dogs in Utility Dog obedience competition has shown that this procedure is adequate to ensure that the dog's choice will not be influenced by the articles' previous scenting histories.

All three dogs had been trained and were handled in all tests by I.L.B. 'Stranger' scent was imparted by any other individual, regardless of sex or degree of familarity to the dog. In some cases the experimenter imparted the stranger scent, but this was done by a third party whenever possible. Again, experience in Utility Dog obedience competition has shown that the ability of dogs to perform scent discrimination tests of this type is not related to either the sex or the degree of familiarity of the dog with the individual imparting the unfamiliar scent.

Experimental Design

Each experimental series consisted of 14–34 replications (usually 20) of given choice pairs, half using metal articles and half using leather articles. The choice pairs and the rationale behind them were as follows.

(1) Handler's hand versus no human scent. This experiment simply ensured that all the subjects had intact olfactory senses, and were appropriately trained.

(2) Handler's elbow versus no human scent. This experiment verified that the antecubital space indeed provided sufficient detectable odours.

(3) Handler's hand versus stranger's hand. This experiment verified that these dogs were capable and willing at the time of each test to perform the Utility Dog task for which they had been trained.

(4) Handler's elbow versus stranger's hand. This was the most important experiment in the evaluation of the individual odour identification hypothesis. If the hypothesis holds, the dogs should still be able to discriminate between individuals in this test. If it does not, they should not.

(5) Handler's elbow versus handler's hand. In this experiment we sought to determine whether

there was a body-part specific odour component, that the dogs would choose because of their training familiarity with the hand.

(6) Handler's elbow versus stranger's elbow. In this experiment we sought to determine whether after controlling for the scent stimulus from a previously unknown body part, there was still a detectable individual odour of the handler. Only dog BH was used in this experiment.

In most of the tests, dogs were presented with one replication each of tests (1), (3), (5) and (4) in that order in a single day, and this protocol was continued until the total number of replicates had been performed with each dog. Examining the results of these tests suggested the importance of test (2), which was then performed four times per day with each dog until the requisite number of replicates had been completed. In one experimental series, BH was given test (6) two to four times per day until 20 replicates had been completed, and later, the same subject was given 18 more replications (two per day) of a series consisting of tests (1), (3), (6) and (5) in that order.

Statistical Analysis

Standard statistical tests assume that observations are an unbiased sample of all possible observations (e.g. Sokal & Rohlf 1981). That is, observations are assumed to be independent. In treating each of our repeated samples within an animal as an independent observation we assumed, that results of trials were not affected by the results of previous trials.

This assumption was examined two ways. First, each series was arbitrarily divided into three temporal categories (early, intermediate and late) of approximately equal size and a chi-squared test was performed on the 3 × 2 contingency table (Everitt 1977). If there are 'training' effects observed over the time course of the trials, that is, if performance level changed significantly, each trial could not be treated as an independent entity. Furthermore, Wald-Wolfowitz runs tests (Wilkinson 1988) were performed on each experimental series to search for non-random response patterns even in the absence of training effects.

The replicated single classification goodness-of-fit tests that we used were: (1) binomial test, if the sample sizes were less than 50; and (2) *G*-test, if the sample size was greater than 50 (Sokal & Rohlf 1981). For *G*-tests, results from all subjects were pooled ($G_p = G$ for pooled data) if no statistical heterogeneity between subjects was detected (G_h at $P < 0{\cdot}05$). For contingency tables, *G*-test analysis was used with Yates' correction for sample sizes under 50, and chi-squared analyses were used for larger sample sizes.

Table II. Summary of the experimental results (number of choices: first alternative/second alternative)

Experiment	Dog LR	 AST	 BH
1. Handler's hand/ no human scent	20/0	14/2	20/2
2. Handler's elbow/ no human scent	16/2	—*	16/4 14/2†
3. Handler's hand/ stranger's hand	14/6	11/5	29/5 16/10†
4. Handler's elbow/ stranger's hand	14/6	8/6	11/12 7/9†
5. Handler's hand/ handler's elbow	13/7	13/2	17/4
6. Handler's elbow/ stranger's elbow	—*	—*	9/1 10/6†

*No tests given.
†Results from 1989.

Because of the large number of statistical tests performed, we used a sequential Bonferroni correction for multiple tests (Rice 1989). This correction is more powerful than the standard Bonferroni technique (Miller 1981), and does not require independence of component tests.

RESULTS

There were no significant differences between dogs' scores using metal and leather dumbbells, therefore scores were combined across dumbbell types for all of the following analyses. Results of all tests are summarized in Table II. Furthermore, no training effects or non-random dispersion in the data were detected (at $P < 0{\cdot}05$), therefore each trial was assumed to be an independent sample within these three dogs.

Ability to Distinguish Hands and Elbows from No Human Scent

Dogs had no difficulty distinguishing the scent of their handler's hand from no human scent at all. Of the 58 total trials completed, dogs chose correctly 54 times (93·1%), which is significantly better than random ($G_p = 51{\cdot}294$, $df = 1$, $P < 0{\cdot}001$). The most successful dog (LR) chose correctly on all 20 trials, the least successful (AST) was correct on 87·5% (14/16) of his trials. There was no significant difference between dogs ($G_h = 3{\cdot}65$, $df = 2$, $P > 0{\cdot}5$).

Scent from the handler's elbow was also easily distinguished from no scent. Only two of our three dogs (LR and BH) were used in this test, but they successfully identified their handler's elbow scent in 32 of 38 trials (84·2%), a score statistically indistinguishable from their success rate with hands ($G = 2{\cdot}196$, $df = 1$, $P > 0{\cdot}5$), and also significantly better than random (one-tailed binomial test, $P < 0{\cdot}001$).

Ability to identify handler's hand

According to their training, the dogs were assumed to be able to distinguish the scent of their handler's hand from that of a stranger. The dogs were clearly capable of this task, as they chose correctly in 75·7% (53/70) of the trials ($G_p = 19{\cdot}431$, $df = 1$, $P < 0{\cdot}01$) and the dogs were statistically indistinguishable ($G_h = 3{\cdot}442$, $df = 2$, $P > 0{\cdot}5$).

Ability to distinguish stranger's hand from handler's elbow

Could these same dogs identify their handlers when presented with the scent of his elbow compared with the scent of a stranger's hand as the individual odour theory would predict? Overall only 57·9% (33/57) of the trials resulted in correct identification of handlers' dumbbells, a result not statistically different from random ($G_p = 1{\cdot}427$, $df = 1$, $P > 0{\cdot}5$; $G_h = 2{\cdot}194$, $df = 2$, $P = 0{\cdot}5$).

Body-part-specific versus individual-specific cues

If human scent is perceived by dogs as having 'hand-specific' (or body-part specific) scent components plus 'individual-specific' components, the dogs' training regime by teaching them to identify handler's-hand-specific scents, may have resulted in confusion when hand-specific and individual-specific cues were in conflict as they would be in the above experiment. To address this possible confusion, we uncoupled 'hand' and 'individual' information by analysing choices between handler's hand and handler's elbow, as well as between handler's elbow and stranger's elbow.

When given a choice between the scent of their handler's hand and elbow, the dogs chose their handler's hand scent 76·8% ($N = 56$) of the time ($G_p = 16{\cdot}945$, $df = 1$, $P < 0{\cdot}01$), suggesting that there indeed is a distinctive hand scent that dogs had inadvertently been trained to identify. The performance of the three dogs was statistically indistinguishable ($G_h = 2{\cdot}559$, $df = 2$, $P > 0{\cdot}5$).

If there are individual odours as well as body part odours, dogs might be expected to be able to choose scent of their handler's elbow when compared with that of a stranger's elbow. We decided to perform these experiments only after the initial results had been analysed, and only one dog (BH) was still available for the test. One series of tests was run late in the first experimental series and a second series was run in 1989. In the latter test series, we first verified that BH could still correctly identify the scent from his handler's hand as opposed to that from a stranger's hand (100% correct in 16 trials), and still could not choose better than randomly between the scents of his handler's elbow and a stranger's hand (chose handler's elbow in 43·7% of the 16 trials) before conducting the tests of handler's elbow scent versus stranger's elbow scent. In all of the elbow tests combined ($N = 26$), BH chose the scent from his handler's elbow 73·1% of the time (one-tailed binomial test, $P = 0{\cdot}08$), nearly the same response that all dogs showed towards the scent of their handlers' hands although because of the smaller sample size the result is no longer statistically significant. However, these results suggest that there may be an identifiable individual scent in addition to the body-part specific scent, but that without special training, the discrimination of scents from different individuals will succeed only when scents are obtained from the same body part of each subject.

DISCUSSION

These experiments show clearly that properly trained dogs are capable of distinguishing between the hand scents of individual humans. They also show that even under the best of circumstances, individual identification is less than perfect. Our most extreme result was that dogs distinguished objects with scent from their handler's hand 93% of the time from objects with no human scent.

Consistent identification of an individual disappeared, however, when the dogs were asked to compare scents from separate parts of the body. There are two possible explanations for this result. The most obvious is that individual human odours are not detectable by dogs. Individual odours for specific body parts may exist, but there is no general scent that dogs can match with individuals. If this interpretation is correct, the forensic ramifications would be enormous. Quite simply, it would mean that dogs cannot be used to match humans with physical evidence.

An alternative interpretation is that individual human odours exist and are detectable by dogs. However, because there are also distinctive body-part odours, training methods that rely on matching odours from only a single body part are inappropriate for matching odours from other body parts. Indeed our experiments suggest that this may be the case. If so, then alterations in training regimes may allow appropriately trained canines to detect individual scents regardless of the body parts from which the scent was obtained.

If the dogs were confused by the conflicting signals inherent in distinguishing between the scent from the handler's elbow and that from a stranger's hand, we might expect them to show more indecision in such tests than in the more straightforward case when they had to choose between odours from a stranger's hand and their handler's hand. We can estimate whether such relative indecision occurred by considering the number of times each item was sniffed and/or the time required to make a decision. Indeed, by this measure, two of the three dogs (LR and AST) showed greater indecision when choosing between the scent from the handler's elbow and that from a stranger's hand (one-tailed $t=1{\cdot}792$, $df=38$, $P=0{\cdot}041$ and $t=1{\cdot}803$, $df=28$, $P=0{\cdot}041$, respectively), although the third dog (BH) did not (one-tailed $t=0{\cdot}884$, $df=56$, $P=0{\cdot}154$).

The idea of both individual and regionally specific body odours is not new. Löhner (1927) demonstrated that human beings can distinguish in other humans both regional and individual odours. In controlled experiments, he determined that humans perceived greater differences between regional odours of the same person than individual differences between two people. Such results are in complete accord with the findings reported here.

The identification of individuals on the basis of information that the dog obtains from scent articles is an important component of the use of dogs in law-enforcement activities and has been often accepted as evidence in court proceedings (Lowe 1981; Tolhurst & Reed 1984). In some cases, such identification has been used as the sole source of evidence upon which to argue for the death penalty in criminal prosecution for murder (e.g. State of Florida versus Eugene Wiley, Jr; Circuit Court of the 18th Judicial Circuit, Brevard County). These legal aspects of scent article identification represent a clear practical application of the results we report in this study. In particular, our results suggest that such identifications may not be reliable unless the articles have been scented from the same part of the body from which the dog had been previously trained to make discriminations. Particularly open to question, for example, would be the use of a piece of clothing from the upper part of the body (e.g. hat, gloves, shirt) as a reference scent article on the basis of which the dog is expected to select a track of human scent made by footsteps through the environment. Although it may indeed be possible to train individual dogs to perform such tasks, the results presented here make it clearly incumbent upon the individual dog trainer or handler to demonstrate that his or her dog can indeed perform the required scent identification tasks with an acceptable degree of statistical reliability, before evidence based on the performance of such a dog should be accepted in a court of law.

Finally, the use of trained animals has been somewhat overlooked in the biological community as a tool for examining the information content of animal signals. We suggest that such animals might be used to ask a host of questions, ranging from issues in the functional rate of environmental signal degradation to the complex details of kin recognition.

ACKNOWLEDGMENTS

Much of the early testing reported in this study was undertaken as part of class exercises and with the assistance of students and instructors in the animal behaviour courses at the University of Georgia, Augusta College and the Woods Hole Marine Biology Laboratory. I.L.B. is particularly grateful to Jelle Atema for early help and encouragement. Additional advice and insight have been obtained from discussions with James Dearinger, Sgt Thomas A. Knott (ret.), and W. Herbert Morrison and Phil Hoelcher. We are grateful to the many people, particularly Michael Newman and Howard

Zippler, who assisted in conducting these tests and provided us with abundant 'stranger' scent. Laura Brooks was particularly helpful in assisting with the 1989 experiments. Support promoting collaboration between the authors was provided by a travel contract from the Oak Ridge Associated Universities and by the Henry Rosovsky Fund to Support Junior Faculty in the Sciences at Harvard University. Manuscript preparation was supported in part under a contract (AC09-76SR00-819) between the U.S. Department of Energy and the University of Georgia.

REFERENCES

American Kennel Club. 1979. *The Complete Dog Book: Official Publication of the American Kennel Club.* New York: Howell Book House.

Bauman, D. L. 1986. *Beyond Basic Dog Training.* New York: Howell Book House.

Budgett, H. M. 1933. *Hunting by Scent.* New York: Charles Scribner's Sons.

Everitt, B. S. 1977. *The Analysis of Contingency Tables.* London: Chapman & Hall.

Löhner, L. 1927. Über Fremdidentifizierung und Anzeigen von Witterungsübereinstimmung. *Z. Vereins dtsch. Schäferhunde (SV) Sit München*, **26,** 566–572.

Lowe, B. 1981. *Hunting the Clean Boot: the Working Bloodhound.* Poole: Blandford Press.

Miller, R. G. Jr 1981. *Simultaneous Statistical Inference.* New York: McGraw Hill.

Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution*, **43,** 223–225.

Sokal, R. R. & Rohlf, F. J. 1981. *Biometry.* 2nd edn. San Francisco: W. H. Freeman.

Tolhurst, W. D. & Reed, L. F. 1984. *Manhunters: Hounds of the Big T.* Puyallup, Washington: Hound Dog Press.

Wilkinson, L. 1988. *Systat: the System for Statistics.* Evanston, Illinois: SYSTAT.

APPENDIX E

ELSEVIER

Forensic Science International
69 (1994) 111-118

Forensic Science International

The ability of dogs to recognize and cross-match human odours

G.A.A. Schoon[a]*, J.C. De Bruin[b]

[a] *University of Leiden, Department of Criminalistics and Forensic Science/Institute of Evolutionary and Ecological Sciences, Ethology Group/National Dutch Police Tracker Dog Centre, PO Box 9520, 2300 RA Leiden, The Netherlands*

[b] *Rotterdam Police Dog Training Centre, PO Box 70023, 3000 LD Rotterdam, The Netherlands*

Received 6 April 1994; accepted 14 June 1994

Abstract

The Dutch police use specially trained dogs, *Canis familiaris*, to identify criminals from scented objects left at the scene of the crime. As part of a study of the reliability of these identifications, experiments were performed to evaluate whether these dogs could match scents collected from different parts of the body. Dutch police dogs are capable of matching hand scent to scent collected in the crook of the elbow and vice versa, and of matching trouser pocket scent to hand scent. It appears that training and familiarity with the person whose scents have to be matched enhance the performance of the dogs. The reliability for judicial purposes entails further research since our knowledge of the sense of smell is limited.

Keywords: Olfactory identification; Scent traces; Dogs

---

1. Introduction

Police use the olfactory acuity of dogs, *Canis familiaris*, to investigate crimes. The dogs are trained to detect narcotics, explosives, or humans and are used in 'scent identification lineups' where the dog identifies a perpetrator on the basis of scent traces left at the scene of the crime. The validity of these scent lineups has been questioned [1] and there is some debate about the capability of dogs to recognize individ-

---

* Corresponding author.

SSDI 0379-0738(94)01565-M

ual human odours. The experiments of Brisbin and Austad [2] showed that when trained dogs were faced with the choice between the hand-scents of master and stranger, they chose their master's scent but faced with the choice between the hand-scent of a stranger and the crook elbow-scent of their master, they no longer preferred the master's scent, although this was expected. The authors concluded that either 'there is no such thing as an individual human odour', or that the dogs they used 'did not spontaneously identify individual odour components of scents taken from different parts of the body'.

These two possible explanations of their findings have been the subject of study by several people in the past. On the point of individual odours, evidence of individual, genetically controlled odours exists [3] and that dogs are capable of discriminating between these [4–7]. However, there is also evidence that different body parts have different odours due to local (scent) glands and varying 'regional' conditions for bacteria. To humans the similarity between 'regional' odours of different individuals is greater than different regional odours of a single individual [8]. The second explanation, that the dogs were not able to match smells collected from different body parts, might well be a matter of training. If different body parts have (partly) different odours, a dog must be trained to focus on the common elements.

The police in The Netherlands use the ability of trained dogs to discriminate between human beings on the basis of odour to identify perpetrators through a match-to-sample experiment, matching the odour of a suspect to scent traces from the scene of the crime. Positive identification by a dog can lead to suspects being held in detention prior to trial for further investigation, or can be used by a judge as part of the evidence on which he rules. The difference in training between the American and Dutch dogs is important to emphasize: the American dogs were always confronted with hand-odour only, and were trained to retrieve articles scented by the hand of their master only, but the Dutch dogs are trained with a variety of scent sources belonging to many individuals in match-to-sample trials directing the dog to focus on the individual human element. If training is an issue, the Dutch police dogs should do better at identification tasks involving different regional scents than the American dogs. We undertook a similar series of experiments to those of Brisbin and Austad [2], in which we sought to establish whether the Dutch police dogs were able to match odours collected on different parts of the body.

2. Material and methods

2.1. Experimental subjects

The experiments were done in 1992 and in the first quarter of 1993 with six Dutch police dogs that had been trained to perform scent identification tasks described below and had been certified in 1983 or later (in the 'speurhond menselijke geur' examination, which means human scent tracker dog). The dogs varied in age between three and eight years; two of the six dogs were male, and they were all Shepherd dogs (German, Malinois or mixed). The participation of these dogs was not equal owing to different workloads, illness and death, and varied in the range 9–45 trials.

The examination and regulations surrounding it had been modified in 1991, and two dogs had passed the 'old' exam, four the 'new'. The following description of the examination is limited to the scent identification task which had to be performed correctly. All preparations were carried out by an examiner to prevent the dog from picking up any other cue than scent (so-called 'Kluge Hans' errors). For the old exam, six people were asked to wash their hands with non-perfumed soap and to dry them on paper towels. After 5 min, each was asked to hold two stainless steel tubes in their hands for 5–10 min. The tubes were then collected separately in glass bottles or in plastic bags, and handled with tongs thereafter. One of these six people was considered 'suspect' and he was also asked to scent a sample object by putting it in his pocket or holding it in his hand. The examination was usually performed outside on a paved surface. The tubes were placed in two rows by the examiner. In this 'scent-lineup' each row contained the six different scents, the position of the suspect's tube was random and unknown to the handler. After these preparations, the handler would come to the area and let his dog sniff ('take air') at the sample object scented by the suspect by holding the object in one hand and directing the nose of his dog to the object with his other hand. He would then command his dog to bring him the corresponding tube in the first row. The dog was left free to make his choice. After a correct retrieval this procedure was repeated for the second row, and if the corresponding tube was retrieved again the dog had passed this part of the exam.

The scent-identification task in the new examination was for a large part the same, except that the two rows were not identical but contained 12 different scents. In one row a corresponding tube had to be retrieved, the other row was a 'control' row in which the dog could not find a corresponding tube and had to refrain from picking up one. The handler did not know which of the two rows contained the corresponding tube, nor the position of the tube in that row. The dog had to perform correctly in both rows to pass.

At the beginning of the training, the old examination protocol is followed for the scent identification tasks, and the sample objects are hand-scented metal objects. Gradually the sample objects are varied in material, age, and method of scenting (scenting objects in pockets predominates). The new examination protocol is introduced later and practised less frequently. In the beginning, 'suspects' and extra scents usually belong to familiar persons, and scents of strangers are introduced later. After their examination, the handlers keep their dogs in training and also perform forensic identification tasks. In 1991 yearly examination became compulsory.

2.2. Experimental protocol

The experimental protocol followed the normal working routine. The 10 cm stainless steel tubes were cleaned with soap, rinsed and then boiled in tap water for ≥30 min, and were handled with tongs thereafter. The people scenting the tubes worked in the building of the Police Dog Training Centre in Rotterdam where the experiments took place (these people were familiar to two of the participating dogs), or were visitors to that Centre. They did not wash their hands prior to scenting the tubes since this is normal in training. People scenting the tubes by hand held them for 3 min and then replaced them in sterilized glass jars with twist-off tops. Tubes scented in the pocket or in the crook of the elbow during the 3 min were handled with tongs to prevent contamination and were put in separate sterilized glass jars. Most of the people that scented tubes were male, the 'suspects' were always male. The tubes were usually prepared on the same day that the experiments were done, but, occasionally, the day before. This is also done in training, during examinations and in normal police procedures.

All trials were performed at the Rotterdam Police Dog Training Centre. The working area was an indoor area of 10×10 m^2 with a smooth floor that was cleaned daily. A trial set consisted of three trials following normal protocol. Six tubes, each having been held by a different person, were laid down in a row. The position of the 'suspect' was random. The handler then entered with his dog and let it 'take air' of the sample object, in this case another steel tube that had been prepared by the 'suspect', and ordered to retrieve the corresponding tube. After a retrieval, or when the handler decided that the dog was not able to make a match (usually after the dog had sniffed each tube twice), the handler would leave the room with the dog. All tubes were removed and a new row would be prepared, taking care to lay down the fresh tubes on a new area on the floor. The second and third trials of the set followed the same protocol as above. The handler did not know the position of the correct tube in advance but was informed of the result between trials. All proceedings were videotaped.

2.3. Experimental design

The experiments were designed to check the results obtained by Brisbin and Austad in 1991 [2]. Each trial set consisted of three matches that the dog had to make, as follows.

(1) Matching elbow to hand scent: the sample tube that the dog was given to smell had been scented in the crook of the elbow, and the six tubes in the row had been scented by hand. This had never been practised by the dogs.
(2) Matching hand to elbow scent: the sample tube had been scented by hand, the matching tube (and two others) in the row had been scented in the crook of the elbow, and the other three tubes by hand. By mixing hand and elbow scents in the row we could examine preference, as this was a crucial element in Brisbin and Austad's experiments [2]. Retrieving a crook-of-the-elbow-scented tube was new for the dogs.
(3) Matching pocket to hand scent: the sample tube had been scented in the pocket, and the six tubes in the row had been scented by hand. This method is predominantly used in training.

The dogs were confronted with one trial set on a day, and the days on which these experiments were done were spread over a period of 15 months. The trials within a set were given in a quasi-random way (Table 1), often starting with match 1 and ending with 3. Within a single trial set, the odours used belonged to the same set of six people. To preclude memory effects, the number of trials in a set was limited to three.

Table 1
Distribution of trials within the 59 trial sets

Order in trial set	1st	2nd	3rd
Elbow ⇒ hand	31	14	14
Hand ⇒ elbow	16	26	17
Pocket ⇒ hand	12	19	28

A 'correct' choice was noted when the dog picked up the matching tube and brought it to its handler. A 'wrong' choice was noted when the dog picked up another tube and brought it, and a 'no choice' was noted when the dog had smelled at all of the tubes (usually twice) but did not pick up any, after which he was recalled by the handler. The 'no choice' option was essentially a decision made by the handler, based on his experience with the dog.

2.4. Statistics

Chi-square testing was used throughout. Where the number of observations was too small for testing, data were grouped into larger categories (correct-not correct, wrong-not wrong). Since the dog could choose between six tubes and a match was potentially possible in all the trials, the chance level was 16.7% for each trial.

3. Results

3.1. Differences in performance between dogs

In regular police practice, it is the result of the identification test that counts, regardless of which individual dog made the identification. The assumption is that, having passed the test, the results of all dogs are equal. We tested the results of the three dogs that contributed the most to the test series and found that the performance of one of the dogs was significantly better ($P < 0.01$) than the other two. However, the results of the tests are handled as a whole hereafter. We feel that this is legitimate since the conclusions that are drawn are reflected in the results of each of the dogs separately, where there were enough data to test this.

3.2. Differences between trial types

The number of correct choices was significantly better ($P < 0.01$) than chance (16.7% of 59 = 9.8 trials) in all trials (Table 2). The performance of the dogs in the first trial of the sets did not differ significantly from the overall performance. The pocket ⇒ hand trials were performed significantly better ($P < 0.025$) than the other two trials types. There was no difference in performance between the hand ⇒ elbow or elbow ⇒ hand trials. In the hand ⇒ elbow trials the dogs had to retrieve a tube that was scented in the crook of the elbow. Analysing the incorrect choices showed that the dogs picked up hand-scented tubes versus crook-of-the-elbow-scented tubes ac-

Table 2
Combined results of six dogs

	N correct	*N* no choice	*N* wrong
Elbow ⇒ hand	19 (32%)	16	24
Hand ⇒ elbow	19 (32%)	15	25
Pocket ⇒ hand	34 (58%)	6	19

cording to their proportion in the row (14 vs. 10 mistakes, 3 vs. 2 proportion in the row).

3.3 Familiarity with scents influences the results

The dog that performed significantly better was based at the Rotterdam Police Dog Training Centre where the experiments were prepared and conducted. Reviewing its results, we found a marked difference in results between tests where the 'suspect' was a person working at the Centre and those when the 'suspect' was a visitor. When the dog knew the 'suspect', 73% of the trials were correct, when the 'suspect' was a familiar scent (often used in training situations) 67% were correct, and when the 'suspect' was a complete stranger to the dog 25% of the trials were correct, but the numbers are too small to test for statistical significance.

4. Discussion

The evidence of our experiments lead to the conclusion that the Dutch police dogs are capable of cross-matching scents collected from different body parts. We are led to hypothesize that this capability is a result of the particular training of these dogs, for the following reasons.

(1) The trial type that was practised most, pocket ⇒ hand, was performed best. One can argue that since people often place their hands in their pockets, pockets contain a lot of 'hand odour' and that this could facilitate the choice for the dog, but on the other hand trouser pockets are close to the groin which is one of the human body areas most richly endowed with scent-producing glands. The tubes that were scented in the pocket for these experiments were not directly touched by hand but picked up the mixed scent in the pocket.
(2) Although the dogs had not been particularly trained on crook-of-the-elbow scents, the dogs performed better than chance. There seems to be enough similarity in scent between tubes scented by hand and those scented in the crook of the arm for these dogs to make the connection, but apparently there is enough difference between them to lead to a larger number of mistakes.
(3) Although the dogs have always been trained to collect tubes that have been scented by hand or in pockets, they show neither aversion to, nor preference for, tubes scented in the crook of the elbow.

The inability of the American dogs in Brisbin and Austad [2] to recognize their masters crook-of-the-elbow scent can be explained by their lack of specific training. Since the dogs were only confronted with hand-held objects, and were trained to only retrieve objects scented by the hand of their master, the choice between a hand-scent of a stranger and an elbow-scent of their master is an ambiguous one. There is no correct choice, i.e. no master's hand. The dogs had never been trained to focus on the individual personal smell of their master, and two of the three dogs used by Brisbin and Austad showed greater indecision (taking more time for their choice) when faced with this ambiguous question.

The importance of training and experimental design on the capability of dogs to recognize personal human odours has been stressed by other investigators, as summarized by McCartney [8]. He refers to work done by the Menzels in the 1930s, who concluded that dogs are able to relate regional body odours to individual odours, but cautioned that many investigators who found that dogs were not able to do this made the fundamental mistake of supposing that dogs are 'naturally' interested in human individual scent.

Training can have profound effects on the results of experiments in many ways, and caution should be taken in interpreting results. The good results of the 'best' dog in our study were obtained in experiments using the scents of familiar people. This factor is common knowledge to dog trainers and is used in dog training: many manuals advise to begin training with the scent of the master, to then move to the scent of familiar persons and finally to strangers. Dogs show different olfactory responses to known and unknown people [9] and can be trained to recognize their masters scent under extreme conditions [10]. For forensic purposes dogs should be trained with scents of strangers. The most common forensic sample objects are tools and weapons, so training on hand-held objects seems logical. A forced-choice method can obviously not be used in forensic applications.

Dogs have remarkable olfactory abilities that can be of great use in police investigations. Dogs are able to identify human scent on glass slides that were lightly fingerprinted, and then kept indoors for 1 month, or outdoors for ≤2 weeks [11]. Other experiments done as part of our investigations have shown that dogs can match steel tubes when the sample object (steel tube) was scented by hand for no more than 5 s, selecting the corresponding scent from a choice of six. However, great care has to be taken in training and experimental design, and in the light of forensic application special attention has to be paid to the number of mistakes a dog makes. Reviewing dog scent lineups, Taslitz [1] concluded that the faith judiciary powers place on results of such lineups is not based on scientific proof. Our experiments show that while dogs are capable of performing scent lineups, in the simple experimental setup of a choice between six, a large number of mistakes are made. However, the new identification task and the yearly examination improve on this, therefore the level of performance found in this study cannot be extrapolated.

5. Acknowledgements

This study was part of a research project funded by the Ministry of Justice. We would like to thank the National Dutch Police Tracker Dog School in The Hague for their general cooperation and all handlers for performing the trials. The discussions with Prof. dr. P. Sevenster were extremely helpful and we thank him for his advice and comments throughout the experiment. We also thank Prof. dr. E.R. Groeneveld and Prof. dr. A.D.F. Addink for their comments on the manuscript.

References

[1] A.E. Taslitz, Does the cold nose know? The unscientific myth of the dog scent lineup. *Hastings Law J.*, 42 (1990) nr.1 15–134.

[2] I.L. Brisbin Jr and S.N. Austad, Testing the individual odour theory of canine olfaction. *Anim. Behav.*, 42 (1991) 63–69.

[3] D.M. Stoddart, *The Scented Ape*, Cambridge University Press, Cambridge, 1990.

[4] H. Kalmus, The discrimination by the nose of the dog of individual human odours and in particular of the odour of twins. *Br. J. Anim. Behav.*, 3 (1955) 25–31.

[5] J. Szinak, Identification of odours. *Int. Crimin. Police Rev.*, (1985) 58–63.

[6] P.J. Hepper, The discrimination of human odour by the dog. *Perception*, 17 (1988) 549–554.

[7] J.C. de Bruin, *The Detection Dog and Science*. Report from Dog Section, Rotterdam Municipal Police, 1989.

[8] W. McCartney, *Olfaction and Odours*, Springer-Verlag, Berlin, 1968.

[9] J.C. Filiatre, J.L. Millot, and A. Eckerlin, Behavioural variability of olfactory exploration of the pet dog in relation to human adults. *Appl. Anim. Behav. Sci.*, 30 (1991) 341–350.

[10] G.J. Romanes, Experiments on the sense of smell in dogs. *Nature*, 36 (1987) 273–274.

[11] J.E. King, R.F. Becker and J.E. Markee, Studies on olfactory discrimination in dogs: (3) ability to detect human odour trace. *Anim. Behav.*, 12 (1964) 311–315.

APPENDIX F

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.*

	Choice of Dog		
Reality ⇓	Odor Suspect	Odor Decoy	No Odor
suspect = perpetrator	+: correct identification	−: wrong	0: miss
suspect ≠ 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 ≠ 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, $p = .03$).

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	+

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

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

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

	Suspect = Perpetrator					Suspect ≠ 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:

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

and

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

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 ≠ 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 ≠ 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 ≠ 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

APPENDIX G

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, November 2009, Vol. 54, No. 6
doi: 10.1111/j.1556-4029.2009.01153.x
Available online at: interscience.wiley.com

Davia T. Hudson,[1] *Ph.D.; Allison M. Curran,*[1] *Ph.D.; and Kenneth G. Furton,*[1] *Ph.D.*

The Stability of Collected Human Scent Under Various Environmental Conditions*,†

ABSTRACT: Human scent evidence collected from objects at a crime scene is used for scent discrimination with specially trained canines. Storage of the scent evidence is usually required yet no optimized storage protocol has been determined. Storage containers including glass, polyethylene, and aluminized pouches were evaluated to determine the optimal medium for storing human scent evidence of which glass was determined to be the optimal storage matrix. Hand odor samples were collected on three different sorbent materials, sealed in glass vials and subjected to different storage environments including room temperature, −80°C conditions, dark storage, and UVA/UVB light exposure over a 7-week period. Volatile organic compounds (VOCs) in the headspace of the samples were extracted and identified using solid-phase micro-extraction–gas chromatography/mass spectrometry (SPME–GC/MS). Three-dimensional covariance mapping showed that glass containers subjected to minimal UVA/UVB light exposure provide the most stable environment for stored human scent samples.

KEYWORDS: forensic science, forensic chemistry, human scent, volatile organic compounds, solid-phase micro-extraction, gas chromatography–mass spectrometry, storage

For over a century, canines have been successfully used in human scent identification in many European countries, in particular, the Netherlands, Poland, Belgium, Germany, and Hungary. Human scent identification line-ups establish an association between a suspect and an object or location based on canines matching human scent collected from a crime scene to scent collected from the hands of a suspect. This identification is based on the theory that every human has a unique odor and canines have the ability to discriminate between these odors (1).

The scent identification line-up is a controversial type of dog scent evidence presented in courts of law (2). Scent identification line-ups represent a relatively new evidentiary tool in the United States. The introduction of human scent evidence has been challenged in court due to the limited scientific research in this field (3–5). Due to the variability with which scent evidence is collected and analyzed across different agencies, such evidence comes under much scrutiny (4,5). For this evidence to be accepted in a United States court of law, it must satisfy the Kelly-Frye, Daubert, or Federal Rules of Evidence depending on if it is a federal offence, the crime committed, and the state in which the case is being tried. In a recent US court, *People of the State of California versus Benigo Salcido*, human scent evidence evaluated by canines was challenged. Some of the issues raised included the uniqueness of human scent, survivability of human scent, and whether canines can be trained to discriminate between scents (6). Numerous testimonies were presented by expert witnesses resulting in the court ruling that human scent evidence can be admissible if: "the person performing the technique used the correct scientific procedures, the training and experience of the dog and dog handler prove them to be proficient, and the methods used by the dog handler in the case are reliable" (6). This case demonstrates the need for the use of robust scientific procedures to produce reliable, reproducible scent evidence that will be admissible in a United States court of law.

Human scent samples for canine use are usually collected utilizing either a direct collection procedure or an indirect collection procedure. The direct collection method involves collecting an article of evidence from the scene of the crime, whereas the indirect method involves the use of a sorbent material to collect the scent from the article of evidence (7). The sorbent material that is employed is dependent on the protocol of the specific country, although cotton-based sorbents are usually used in Europe (8). A suspect is often not immediately identified so the storage of samples is required. Western European countries are currently storing their human scent samples in rooms which are at a constant temperature and are exposed to little or no daylight (8). In Asia, China has recently reported the development of a "scent bank" where scent samples collected on various sorbents are stored at −18°C (9,10).

Solid-phase micro-extraction–gas chromatography/mass spectrometry (SPME–GC/MS) is an analytical technique which has been used for the extraction of volatile organic compounds (VOCs) which are present in the headspace of various forensic samples such as drugs, explosives, and human scent. SPME–GC/MS has proven to be a viable technique for the extraction, separation, and identification of the compounds which are present in the headspace of scent samples (11–16). The headspace of scent samples collected and aged can be distinguished chromatographically based on a combination of the relative peak area ratios of the common compounds present in these samples. Due to the volatile nature of scent samples, it is important to determine the optimal materials and procedures for the collection and storage of human scent (14). The purpose of this study is to evaluate a variety of storage container types and to determine the effects of various storage conditions on collected human scent samples.

---

[1] Department of Chemistry and Biochemistry, International Forensic Research Institute, Florida International University, Miami, FL 33199.

*Support in the form of a grant was obtained from the Netherlands National Police.

†Presented in a poster form at the 58th Annual Meeting of the American Academy of Forensic Sciences in Seattle, WA, February 20–25, 2006.

Received 30 June 2008; and in revised form 25 Oct. 2008; accepted 26 Oct. 2008.

Materials and Methods

This study was approved by the Florida International University Committee on Human Subjects (Institutional Review Board).

Materials

Sorbent materials used were DUKAL brand, sterile, 2 × 2 inch, 8 ply, gauze pads (DUKAL Corporation, Syosset, NY), Kings Cotton, non-sterile, 2 × 2 inch sorbent material (Seafarma, the Netherlands) and Johnson and Johnson brand, sterile, 2 × 2 inch gauze pads (Johnson and Johnson Consumer Products Company, China). The extraction solvents for the supercritical fluid extraction were supercritical grade carbon dioxide (Air Products, Allentown, PA) and HPLC grade methanol (Fisher Scientific, Pittsburgh, PA). The heat sealer utilized was a Maxi Seal electric heat sealer (Premium Balloon Accessories, Taiwan).

Different types of containers evaluated as possible storage containers for human scent included: Ziploc, Freezer Guard Seal, Pint Size, 7.0 × 5.25 inch (SC Johnson & Sons Inc., Racine, WI). Kapak Heavy Duty SealPAK Pouches, PET//LLDPE, 4.5 mL thick, 6.5 × 8 inch (Kapak Corporation, Minneapolis, MN), Kapak Aluminized Pouches, tri-layer polymer chemistry featuring an aluminum film, 6.5 × 8 inch (Kapak Corporation), polyethylene pouches, 3 × 3 inch, 2 mL thick (Veripak, Atlanta, GA).

The containers used to hold the sorbent materials for storage were 10-mL glass, clear, screw top vials with PTFE/Silicone septa (SUPELCO, Bellefonte, PA). The soap used for hand washing was Natural, Clear Olive Oil Soap from Life of the Party (North Brunswick, NJ). The SPME fibers used for the headspace extractions were 50/30 μm divinylbenzene/carboxen/polydimethylsiloxane (SUPELCO, Bellefonte, PA).

The temperature and the humidity of the storage conditions were monitored using Thermochron I-Buttons (MAXIM, Dallas, TX). Storage containers used were glass aquarium tanks (All Glass Aquarium, WI) enclosed with aluminum foil (Reynolds Consumer Products Richmond, VA). The light source used was a UVA/UVB reptile light (Energy Savers Unlimited, CA). The −80°C freezer used was a VWR brand (Revco Scientific Inc., Asheville, NC).

SFE Pre-Treatment of Gauze

The equipment used was an ISCO Model 260D Syringe Pump with an SFX 2–10 Supercritical Fluid Extractor. The SFE conditions used included direct spiking of 1000 μL of methanol into the 10 mL extractor vessel, 30 min static extraction followed by a 10 min dynamic extraction at 1.5 mL/min and 4500 psi. The vessel was maintained at 130°C (11).

Evaluation of Different Storage Containers

SFE pre-treated Dukal gauze was sealed into five types of storage containers which include: 10-mL glass, clear, screw top vials with PTFE/Silicone septa, Ziploc, Freezer Guard Seal, Pint Size bags, KPAK Heavy Duty SealPAK Pouches, KPAK Aluminized Pouches, and polyethylene pouches. A heat sealer was used to seal both the KPAK Heavy Duty SealPak and Aluminized pouches as well as the polyethylene, whereas the Ziploc, Freezer Guard bags were sealed using the zipper at the top of the bag. These storage containers were evaluated in triplicate at each of the different time intervals, including 1-, 2-, and 5-week periods. At the end of the time periods each piece of gauze was removed from its respective storage material and placed back into its original vial using tweezers previously rinsed with a bleach solution and dried. Each stored gauze pad was then re-evaluated using a SPME–GC/MS method.

Hand Sampling Procedure

Five hand odor samples were collected per day from six subjects. Samples were collected from each subject over four consecutive days resulting in a total of 20 samples per subject. Subjects were required to wash hands and forearms with clear Olive Oil Soap for 30 sec, rinse with water for 2 min, air dry for 4 min, then rub the palms of hands over forearms for 5 min. Subjects then sampled themselves by holding the pre-treated 2 × 2 absorbent material (DUKAL brand gauze pads, Kings Cotton absorbent material, and Johnson and Johnson brand) between the palms of the hands for 10 min. The sample was then placed back inside the 10-mL glass vial and sealed by the subjects. This sampling procedure was previously determined to be a viable collection technique to obtain individual human scent profiles from the hands (8,13,16), and olive oil-based, fragrance free soap has been shown previously not to contain any previously reported human scent compounds (14).

Storage of Scent Samples

The collected hand odor samples were subjected to four different environmental conditions: room temperature, −80°C temperature, dark, and UVA/UVB light. Samples stored at room temperature were allowed to stand in an open environment over the 7-week period. These samples were subjected to 10 h of fluorescent lighting of *c.* 300–500 lux and 14 h of darkness. The room temperature was controlled to within ± 1°C with an average temperature of 20°C and an average relative humidity of 56 ± 6%. Samples stored at −80°C were maintained at a temperature of −80 ± 2°C. Once removed from this condition for analysis, samples were allowed to equilibrate to ambient condition for 1.5 h before being subjected to a 21 h SPME extraction.

The container used for the dark storage environment was completely enclosed with aluminum foil to prevent the entry of light. The average temperature and relative humidity in this container was 19 ± 4°C and 71 ± 6%, respectively. The container used for storage of the samples subjected to UVA/UVB light was only partially enclosed with aluminum foil with an opening at the top for the positioning of a 500 lux UVA/UVB light source. The 10-mL glass vials which were used for the storage of the scent samples, offer no protection against the transmission of UV light. The samples which were stored in this condition were constantly exposed to the UVA/UVB light source for the duration of the storage period. The average temperature and relative humidity in this container was 22 ± 2°C and 63 ± 3%, respectively.

Environmental controls were prepared by storing each of the three sorbent material types used for collection of hand odor samples in all four environmental conditions and monitored over the time period. The materials were all pre-cleaned using the SFE method which was previously discussed. Four of the five samples collected on each sorbent material were stored in each environmental condition and at the specific time period (week 1, week 3, week 5, and week 7) one was removed and analyzed using SPME–GC/MS (the fifth sample was used for week 0 analysis).

SPME–GC/MS Procedure

The VOCs from the headspace of the vials containing the absorbent material were extracted using 50/30 μm DVB/CAR/PDMS fibers (13). Single headspace extractions of each vial from each of the storage conditions were performed at room temperature for 21 h. The instrumentation used for the separation and analysis of the analytes was an Agilent 6890 GC/5973 MSD with a 0.25 mm × 30 m HP5-MS column which had a 0.25 μm phase film thickness. Helium carrier gas was maintained at a flow rate of 1.0 mL/min. The initial GC oven temperature of 40°C was held for 5 min, followed by a temperature ramp of 10°C per minute to a final temperature of 250°C which was held for 2 min. The mass spectrometer transfer line was maintained at 280°C and the source temperature was 230°C. Mass spectra were repeatedly scanned from 39–300 m/z. Mass spectra data from 2000 to 6600 scans were exported into comma separated values (CSV) format files using the Agilent Chemstation 3D-Export option. The CSV files were transferred into Microsoft Excel (Microsoft Inc.) where matrix manipulations were performed using an in-house written software running on a PC.

Statistical Evaluation

Three-dimensional covariance mapping was used to compare the VOCs present in the week 0 hand odor samples to the VOCs in the aged hand odor samples (week 1, 3, 5, and 7). Utilization of this technique demonstrates whether or not two hand odor samples collected from the same individual remain unchanged over a storage period as it provides an assessment of origin based on pattern recognition and comparison. Comparisons of covariance maps computed from GC/MS data have previously been used to provide a fingerprint for complex samples such as ignitable liquids (17). Three-dimensional covariance mapping was used for the analysis of the data by using mass spectrometry software to export a data matrix comprised of the individual ion abundances for each mass-to-charge ratio for the mass spectra data from scan 2000–6600 of the chromatographic analysis. The covariance matrix is computed by pre-multiplying the exported matrix by its transpose (the rows of the original sample become columns and vice versa). The computed matrix is normalized and two matrices are compared analytically by calculating a distance, D. D is calculated according to the equation below (17):

$$D = \frac{\sum_i \sum_j |z_{N1}(ij) - z_{N2}(ij)|}{2}$$

Z_N represents the covariance matrix which is normalized such that the sum of all matrix elements equal one. The maximum value that can be obtained is 1 and so a similarity index, S, based on D can also be calculated using the equation below:

$$S = 1 - D$$

The similarity index produces values between 0 and 1; 1 demonstrates similarity while a value of 0 shows total dissimilarity.

Discussion

Evaluation of Different Storage Containers

An optimization of storage container type is an important aspect to determining an optimized storage protocol for human scent evidence. Various types of containment were evaluated to determine if the containment matrixes had any contributions of volatile compounds onto gauze materials which were initially determined to be analytically clean at time zero. Figure 1 shows representative chromatograms produced from storage of pre-treated gauze in glass vials, polyethylene pouches, Ziploc Freezer Guard bags, Aluminized Kapak, and Heavy Duty Kapak pouches for the 5-week period. Table 1 displays the average number of overall compounds detected across the triplicate cotton material samples stored in the various containers at week 1, week 2, and week 5 and also displays the number of previously reported human scent compounds detected. The storage container which contributes the least amount of both overall compounds and those previously reported to be components of human scent onto the pre-treated gauze is the

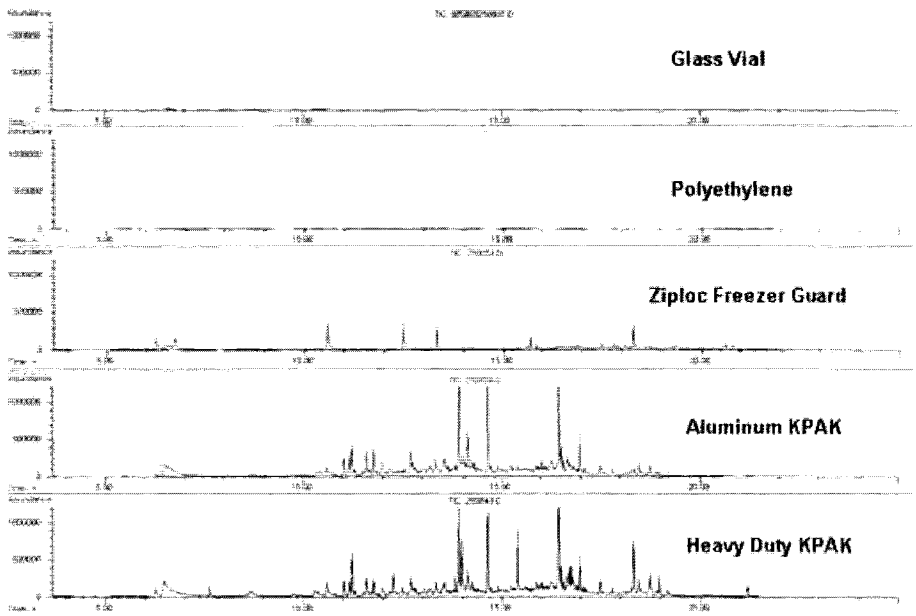

FIG. 1—*Chromatograms showing the effects of storing analytically clean sorbents in various containers for 5 weeks.*

TABLE 1—*Average number of compounds present on analytically clean sorbent materials after storage in different types of containers.*

	Average Number of Compounds			Average Number of Human Compounds		
Storage Container	Week 1	Week 2	Week 5	Week 1	Week 2	Week 5
10-mL glass vial	2	1	5	1	0	1
Polyethylene	19	24	11	6	7	2
Ziplock Freezer Guard	40	19	15	9	3	3
Aluminum KPAK	93	85	88	6	3	2
Heavy Duty KPAK	116	116	107	7	8	6

TABLE 2—*Calculated similarity, S, between hand odor samples collected on different sorbent materials and stored at room temperature over a 7-week period (week 0 vs. weeks 0, 1, 3, 5, and 7).*

Room Temperature			
Time (Weeks)	Dukal	Kings Cotton	Johnson and Johnson
0	1.00	1.00	1.00
1	0.79	0.71	0.74
3	0.59	0.58	0.67
5	0.66	0.52	0.53
7	0.64	0.54	0.49

10-mL glass vial, whereas the material which contributes the most is the Heavy Duty Kapak pouches.

The glass containment evaluated during the storage period displayed the presence of nonane in two of the nine samples, which has been previously reported to be a human odor compound. Four compounds detected in one of the triplicate samples analyzed during week 5 of storage in the glass container were long-chain cyclic alkanes which were present due to SPME fiber degradation after the extended extraction times utilized for analyzing human scent and were not present due to the glass containment. The gauze materials placed inside the polyethylene, Aluminized Kapak, and Heavy Duty Kapak pouches were all sealed utilizing a heat sealer. The compound classes detected in these materials included alkanes, aldehydes, acid-methyl esters, and alcohols. The process of heat sealing may have caused the creation and/or release of many of the compounds detected on the gauze after storage in these matrixes.

In terms of analytical evaluation an increase in the background is a major problem as instruments have limitations in terms of selectivity. Canines, however, depending on their training have demonstrated comparatively enhanced selectivity as they possess the ability to detect targets in the presence of a high background. Although the increase in overall background levels may not pose a problem for canine evaluation of human scent samples the possibility exists that if compounds previously determined to be present in human scent are added to human scent samples through contact with storage materials the scent profile may be altered thus affecting performance of the canines. As glass was determined to be the optimal containment for the gauze collection materials, in the storage environment experiments glass vials were utilized for all subsequent storage of samples.

Storage of Scent Samples

Five hand odor samples were collected per day from six subjects (two subjects per material). Samples were collected from each subject over four consecutive days resulting in a total of 20 samples per subject which were stored in four different environmental conditions. The sorbent materials used were Dukal brand gauze pads, Kings Cotton sorbent material and Johnson and Johnson brand gauze pads. Dukal brand gauze and Kings Cotton are both 100% cotton, whereas the Johnson and Johnson brand gauze is a blend of polyester, rayon, and cellulose.

Environmental controls were monitored across the time period by storing each of the three sorbent material types used for collection of hand odor samples in all four environmental conditions. As previously described in the section above, glass containment provides minimal contribution of previously reported human odor compounds to the stored samples; any detection of nonane was disregarded in this analysis. Storage in the presence of UVA/UVB light did however result in the detection of aldehydes previously reported as being human scent compounds. These compounds were not initially detected in the SFE cleaned gauzes. As this was only observed with the sorbent materials which were stored in the presence of UVA/UVB light, it is being assumed that the UV light may have caused the creation and/or release of the aldehydes detected on the gauze after storage in this condition. The detection of the aldehydes over time was observed mainly on the 100% cotton sorbents. These results suggest that the sorbent materials were being adversely affected by the UVA/UVB light storage. Previous research has shown that materials such as cotton even though they have good resistance to sunlight, degrade with prolonged exposure to ultraviolet light (18).

Room Temperature Storage

Comparisons made utilizing three-dimensional covariance mapping values demonstrated that the scent profiles on all the absorbent materials were changing as the storage period progressed (Table 2). The hand odor samples which were stored on the Dukal brand gauze at room temperature produced a similarity value of 0.64 at the end of the storage period while similarity values of 0.54 and 0.49 were obtained for Kings Cotton and Johnson and Johnson brands, respectively. This showed that Dukal gave the least variation over the 7-week period when compared to the samples stored on the two other sorbent materials. Also, the difference between the similarity values for week 0 and week 1 was greater than the difference between week 5 and week 7. This trend was observed across all three sorbent materials (Table 2). This suggests that the VOCs in the scent samples were changing less as the storage period progressed.

−80°C Storage

Similarity values of 0.64, 0.60, and 0.66 were obtained for Dukal brand, Kings Cotton and Johnson and Johnson brand gauze respectively for the seventh week of storage in −80°C.

Of all the three sorbent materials, Johnson and Johnson brand showed the greatest similarity between the week 0 and the week 7 samples. The Johnson and Johnson gauze also showed a smaller difference between the similarity values for week 5 and 7 when compared to Dukal and Kings Cotton (Table 3).

This shows the 100% cotton materials reacting differently than the Johnson and Johnson gauze in the −80°C storage condition. This can possibly be explained by the characteristic nature of the samples; cotton fibers are hydrophilic and swell in water whereas polyester is hydrophobic and repels water (18,19). Once hand odor samples are collected, it is possible there are small quantities of moisture present on the sorbent material. This could result in the freezing and thawing of the samples during storage and analysis,

TABLE 3—*Calculated similarity, S, between hand odor samples collected on different sorbent materials and stored at −80°C over a 7-week period (week 0 vs. weeks 0, 1, 3, 5, and 7).*

	−80°C		
Time (weeks)	Dukal	Kings Cotton	Johnson and Johnson
0	1.00	1.00	1.00
1	0.65	0.82	0.80
3	0.88	0.92	0.75
5	0.86	0.86	0.72
7	0.64	0.60	0.66

having a greater effect on the 100% cotton sorbent materials more than the Johnson and Johnson brand which is a blend of cotton/rayon and polyester.

Dark Storage

The samples which were stored on Dukal brand gauze in the dark produced a similarity value of 0.67 at week 7 while similarity values of 0.43 and 0.42 were obtained for Kings Cotton and Johnson and Johnson brand, respectively (Table 4). Samples stored in this condition showed a gradual decrease in the similarity values as the storage period progressed. Like the room temperature storage, the differences in the similarity values between the initial weeks (week 0 and 1) were greater than between the final weeks (week 5 and 7) of storage. This trend was observed for all three sorbent materials.

UVA/UVB Light Storage

Hand odor samples subjected to storage in the presence of UVA/UVB light also showed a gradual decrease in the similarity values over the storage period for all absorbent materials investigated (Table 5). The Johnson and Johnson brand gave the greatest change over the 7-week period (7): three-dimensional covariance mapping value of 0.32. Storage in the presence of UVA/UVB light resulted in the detection of methyl esters and aldehydes which were not previously detected in the "fresh" (week 0) hand odor samples. These "new" compounds which were often detected by the third week of storage persisted for the remainder of the storage period. This is similar to what was observed with the environmental controls stored in this condition.

TABLE 4—*Calculated similarity, S, between hand odor samples collected on different sorbent materials and stored in the dark over a 7-week period (week 0 vs. weeks 0, 1, 3, 5, and 7).*

	Dark		
Time (weeks)	Dukal	Kings Cotton	Johnson and Johnson
0	1.00	1.00	1.00
1	0.87	0.53	0.76
3	0.78	0.71	0.70
5	0.72	0.54	0.54
7	0.67	0.43	0.42

TABLE 5—*Calculated similarity, S, between hand odor samples collected on different sorbent materials and stored in the presence of UVA/UVB/ light over a 7-week period (week 0 vs. weeks 0, 1, 3, 5, and 7).*

	UVA/UVB Light		
Time (Weeks)	Dukal	Kings Cotton	Johnson and Johnson
0	1.00	1.00	1.00
1	0.74	0.66	0.58
3	0.71	0.58	0.56
5	0.70	0.71	0.36
7	0.66	0.59	0.32

Also, studies conducted on changes in the lipid composition of fingerprint residue, collected on glass fiber filter paper, have shown that the presence of UV light does produce oxidation reactions resulting in the formation of VOCs such as aldehydes and methyl esters (20). Oxidative degradation of the fatty acid component of sebaceous glands has also been shown to produce aldehydes (21). These are some possible reasons aldehydes and methyl esters were detected but there is no certainty as to whether or not these compounds were created during exposure to UVA/UVB light or they were originally present but not readily released by the sorbent materials. This was however not observed in any of the other storage conditions.

An individual's primary odor compounds have been defined by Curran et al. as the constituents of the odor that are stable over time regardless of diet or environmental conditions (14–16). The compounds which were consistently present in the individual hand odor samples over 4 days of sampling were chosen to be the primary odor compounds and these compounds were monitored over the storage period. The primary odor compounds were determined to be 2-furancarboxaldehyde, phenol, nonanal, and decanal for the hand odor samples collected on the Dukal brand gauze and stored in the presence of UVA/UVB light. The "new compounds" detected after week 3 were benzaldehyde, octanal, undecanal, decanoic acid-methyl ester, and 2-octenal. The hand odor samples collected from a male subject and stored on Kings Cotton in the presence of UVA/UVB light had, as its primary odor compounds, benzyl alcohol, nonanal, decanal, and tetradecane while the "new compounds" detected were benzaldehyde and octanal. For the samples collected from a female subject on Johnson and Johnson brand gauze and stored in UVA/UVB light, the primary odor compounds were found to be nonanal, decanal, undecanal, and dodecanal. Unlike the 100% cotton sorbents, the "new compounds" that were detected and persisted on the Johnson and Johnson brand gauze after week 3 were mainly alkanes such as hexadecane and pentadecane.

Aging Effects

The primary odor compounds only account for a fraction of the overall scent profile (11,14–16). Throughout the storage period, the human VOCs present in the hand odor sample for each of the subjects were monitored via single headspace SPME extractions followed by analysis via GC/MS. Changes in the scent profile whether from the primary odor compounds or additional human compounds in the scent profile were detected by three-dimensional covariance mapping. For all the conditions and sorbent materials monitored, covariance mapping showed that the greatest variation within the scent samples was observed between week 0 and 3 after which the variations between samples decreased (week 3–7) (Tables 2–5). Despite the observed changes in the overall scent profile, the ratios of the monitored primary odor compounds remained consistent (Figs. 2–4).

These results are comparable to an aging study (2 weeks to 6 months) on crime scene objects conducted by Schoon of the Netherlands National Police. The study showed that dogs could faultlessly match odors which were collected on the same day but their performance decreased when instructed to match stored

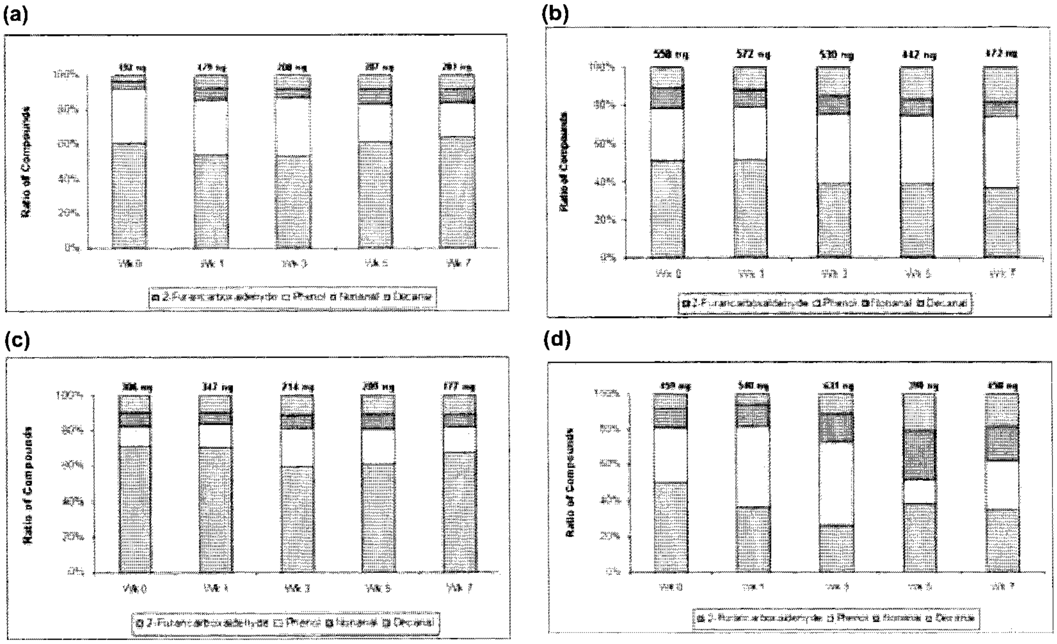

FIG. 2—*Comparison between common VOCs present in hand odor samples collected on Dukal brand gauze from a male subject and stored at (a) room temperature, (b) −80°C, (c) dark, (d) UVA/UVB light.*

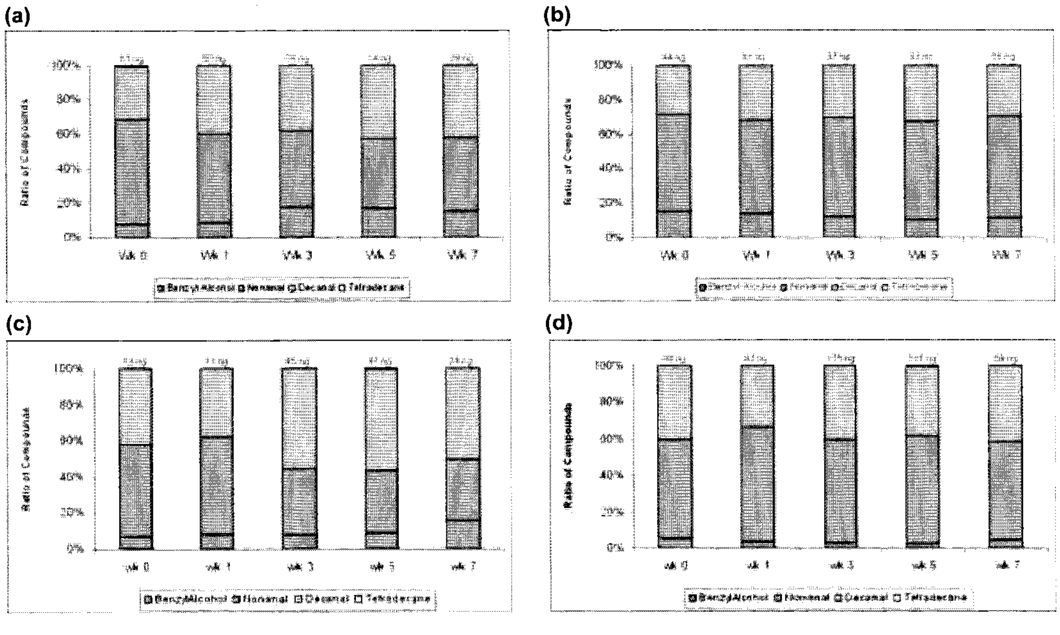

FIG. 3—*Comparison between common VOCs present in hand odor samples collected on Kings Cotton from a female subject and stored at (a) room temperature, (b) −80°C, (c) dark, (d) UVA/UVB light.*

objects to a subject (22). The presence of additional compounds due to storage may mask the primary odor compounds of an individual's scent sample resulting in decreased canine performances when matching aged samples. It is believed that the canines were however still able to make a match as the primary odor compounds are still present in a consistent ratio.

Conclusion

This study of storage containers has demonstrated that when analytically clean cotton materials are stored inside various polymer and aluminized materials a significant amount of compounds are imparted to the cotton material, including compounds previously determined to be present in human scent. Glass has been determined to be the optimal type of storage container for human scent samples as the cotton materials stored in that manner had less overall compounds contributed through the storage method, but also significantly less compounds that have been previously reported to be present in human scent. Glass storage materials are also the storage container used most readily by human scent canine units across Europe (8).

The sorbent materials may have also had an effect on the stored hand odor samples perhaps due to their different chemical compositions. The Dukal brand gauze gave the highest similarity values

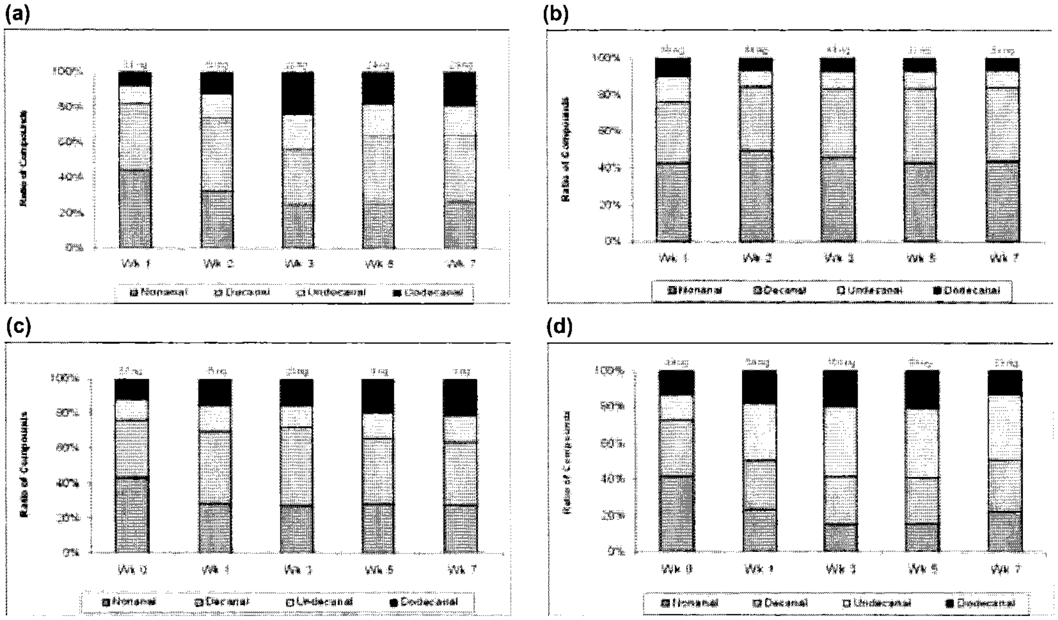

FIG. 4—*Comparison between common VOCs present in hand odor samples collected on Johnson and Johnson brand gauze from a female subject and stored at (a) room temperature, (b) −80°C, (c) dark, (d) UVA/UVB light.*

over the 7-week storage period (7), therefore offering the least variation in all the storage conditions. The Johnson and Johnson brand gauze however, produced the lowest similarity values over the storage period resulting in the most variations for all the conditions. The similarity index also showed a consistent decrease for the Johnson and Johnson compared to the 100% cotton materials. This was likely due to the different fiber chemistries of the sorbent materials. The 100% cotton materials have a more polar backbone and since the majority of the primary odor compounds observed are polar compounds this may have resulted in enhanced collection and retention of compounds compared to the Johnson and Johnson material. The three-dimensional covariance mapping results also showed that the 100% cotton materials did not perform well in the −80°C storage condition as there continued to be great differences in the similarity values in the final weeks of storage (weeks 5 and 7). Whether or not changes using different storage conditions would influence matches by canines was not part of this study and is being evaluated in ongoing studies.

The findings of this study also suggest that for all the environmental conditions studied, the scent profile changed with time with the greatest variations being observed between week 0 and 3 as determined by three-dimensional covariance mapping. The results show that scent samples should not be exposed to excessive amounts of UVA/UVB light as this will result in the detection of a greater number of methyl esters and aldehydes in the headspace of the sample which may alter the human scent profile.

References

1. Stockholm RA, Slavin DL, Kift W. Specialized use of human scent in criminal investigations. Forensic Sci Int 2004;6(3):1–13.
2. Taslitz AE. Does the cold nose know? The unscientific myth of the dog scent lineup. Hastings LJ 1990;42(17):1–179.
3. Mesloh C. Scent as forensic evidence and its relationship to the law enforcement canine [Thesis]. Orlando (FL): Univ. of Central Florida, 2000.
4. *The People, Plaintiff and Respondent vs. Ryan Willis:* Hearing on B160530 Before the Court of Appeal of California, Second Appellate District, Division Two on 115 Cal. App. 4th 379 (2004).
5. *The People, Plaintiff and Respondent vs. Jeffrey Dewayne Mitchell* et al. Hearing on B151026 Before the Court of Appeal of California, Second Appelate District, Division One on 110 Cal. App. 4th 772 (2003).
6. *People of the State of California vs. Benigno Salcido:* Hearing on GA052057 Before the Los Angeles Superior Court on 115 Cal. App. 4th 379 (2005).
7. Eckenrode BA, Ramsey SA, Stockham RA, Van Berkel GJ, Asano KG, Wolf DA. Performance evaluation of the scent transfer unit^{TM} (STU-100) for organic compound collection and release. J Forensic Sci 2006; 51(4):780–9.
8. Schoon A, Haak R. K-9 Suspect discrimination, training and practicing scent identification line-ups. 1st edn. Canada: Detselig Enterprises Ltd, 2002.
9. http://www.english.peopledaily.com.cn/200603/21/eng20060321_252270.html Accessed September 13, 2007.
10. http://www.news.xinhuanet.com/english/2006-03/15/content_4307207.html Accessed September 13, 2007.
11. Curran AM, Rabin SI, Prada PA, Furton KG. Comparison of the volatile organic compounds present in human odor using SPME–GC/MS. J Chem Ecol 2005;31(7):1607–19.
12. Lorenzo N, Wan T, Harper RJ, Hsu YL, Chow M, Rose S, et al. Laboratory and field experiments to identify Canis lupus var. familiaris active odor signature chemicals from drugs, explosives, and humans. Anal Bioanal Chem 2003;376:1212–24.
13. Curran AM, Ramirez CF, Schoon AA, Furton KG. The frequency of occurrence and discriminatory power of compounds found in human scent across a population determined by SPME–GC/MS. J Chromatogr B Analyt Technol Biomed Life Sci 2007;846(1–2):86–97.
14. Curran AM. The analytical determination of the uniqueness and persistence of the volatile components of human scent using optimized collection methods [Dissertation]. Miami (FL): Florida International University, 2005.
15. Schoon GAA, Curran AM, Furton KG. Odor biometrics. In: Li S, editor. Encyclopedia of biometrics. North America: Springer, 2009;1003–8.
16. Curran AM, Prada PA, Furton KG. The differentiation of the volatile organic signatures of individuals through SPME–GC/MS of characteristic human scent compounds. J Forensic Sci. 2010; In press.
17. Sigman ME, Williams MR. Covariance mapping in the analysis of ignitable liquids by gas chromatography mass spectrometry. Anal Chem 2006;78(5):1713–8.
18. Carter ME. Essential fiber chemistry. In: Rubenfeld N, editor. Textile research institute. New Jersey: M. Dekker, 1971;1–53.
19. Bendak A, El-Marsafi SM. Effects of chemical modifications on polyester fibers. Medical Journal of IAS 1991;4(4):275–84.

20. Archer NE, Charles Y, Elliot JA, Jickells S. Changes in the lipid composition of latent fingerprint residue with time after deposition on a surface. Forensic Sci Int 2005;154(2):224–39.
21. Zhang Z, Cai J, Ruan G, Li G. The study of fingerprint characteristics of the emanations from human skin arm using the original sampling method by SPME-GC/MS. J Chromatogr B Analyt Technol Biomed Life Sci 2005;822(1–2):244–52.
22. Schoon GAA. The effect of the ageing of crime scene objects on the results of scent identification line-ups using trained dogs. Forensic Sci Int 2005;147(1):43–7.

Additional information and reprint requests:
Kenneth G. Furton, Ph.D.
Department of Chemistry and Biochemistry
International Forensic Research Institute
Florida International University
Miami, FL 33199
E-mail: furtonk@fiu.edu

APPENDIX H

Anim Cogn
DOI 10.1007/s10071-010-0373-2

ORIGINAL PAPER

Handler beliefs affect scent detection dog outcomes

Lisa Lit · Julie B. Schweitzer · Anita M. Oberbauer

Received: 30 March 2010 / Revised: 13 December 2010 / Accepted: 14 December 2010

Abstract Our aim was to evaluate how human beliefs affect working dog outcomes in an applied environment. We asked whether beliefs of scent detection dog handlers affect team performance and evaluated relative importance of human versus dog influences on handlers' beliefs. Eighteen drug and/or explosive detection dog/handler teams each completed two sets of four brief search scenarios (conditions). Handlers were falsely told that two conditions contained a paper marking scent location (human influence). Two conditions contained decoy scents (food/toy) to encourage dog interest in a false location (dog influence). Conditions were (1) control; (2) paper marker; (3) decoy scent; and (4) paper marker at decoy scent. No conditions contained drug or explosive scent; any alerting response was incorrect. A repeated measures analysis of variance was used with search condition as the independent variable and number of alerts as the dependent variable. Additional nonparametric tests compared human and dog influence. There were 225 incorrect responses, with no differences in mean responses across conditions. Response patterns differed by condition. There were more correct (no alert responses) searches in conditions without markers. Within marked conditions, handlers reported that dogs alerted more at marked locations than other locations. Handlers' beliefs that scent was present potentiated handler identification of detection dog alerts. Human more than dog influences affected alert locations. This confirms that handler beliefs affect outcomes of scent detection dog deployments.

Keywords Dog · Canine · Scent detection · Social cognition · Interspecies communication

Introduction

In the early twentieth century, a horse named Clever Hans was believed to be capable of counting and other mental tasks. The psychologist Oskar Pfungst confirmed that Clever Hans was in fact recognizing and responding to minute, unintentional postural and facial cues of his trainer or individuals in the crowd (Pfungst 1911). The "Clever Hans" effect has become a widely accepted example not only of the involuntary nature of cues provided by onlookers in possession of knowledge unavailable to others, but of the ability of animals to recognize and respond to subtle cues provided by those around them. However, an additional important consideration was the willingness of onlookers to assign a biased interpretation of what they saw according to their expectations.

Experimental paradigms for investigation of animal behaviors are designed to minimize or eliminate confounds arising from the Clever Hans effect. Because the abilities of domestic dogs to respond to human social cues have been extensively documented (reviewed in Miklosi et al. 2007;

L. Lit
Department of Neurology, University of California at Davis,
Sacramento, CA 95817, USA

J. B. Schweitzer
Department of Psychiatry and Behavioral Science,
University of California at Davis, Sacramento, CA 95817, USA

A. M. Oberbauer
Department of Animal Science,
University of California at Davis, Davis, CA 95616, USA

L. Lit (✉)
M.I.N.D. Institute, University of California at Davis,
2805 50th Street, Room 2415, Sacramento, CA 95817, USA
e-mail: llit@ucdavis.edu

Reid 2009), a Clever Hans effect might be particularly prevalent in dogs. Indeed, the reliance of some dogs on human cues has been shown to override olfactory or visual cues indicating the location of food (Szetei et al. 2003). In one experiment, about 50% of dogs would go to an empty bowl indicated by human pointing rather than to a bowl in which the dog had seen and smelled food (Szetei et al. 2003).

This finding was notable in view of the exceptional olfactory acuity in the domestic dog. Humans have capitalized on dogs' olfactory sensitivity through use in an ever-expanding array of scent detection activities (e.g., Horvath et al. 2008; McCulloch et al. 2006; Oesterhelweg et al. 2008; Wasser et al. 2004). Scent detection dogs search an area as directed by their handlers, issuing an operant trained response ("alert") upon detection of their trained scent. However, scent detection dog performance is not solely dependent on olfactory acuity. Cognitive factors such as context dependence (Gazit et al. 2005) and the interaction between training paradigm and the nature of the detection problem (Lit 2009; Lit and Crawford 2006) also can impact performance.

Because the alerting response is initially trained by handler cueing upon dog interest in the desired target scent (e.g., Wasser et al. 2004), it is possible that dogs are also being conditioned to respond to additional unintentional human cues. Generally, trained dogs, including search and rescue dogs, look at humans less than untrained dogs in experimental paradigms requiring dogs to solve a problem such as opening a container (Marshall-Pescini et al. 2009, 2008; Prato-Previde et al. 2008). Indeed, an inverse relationship between owner/handler dependence and problem-solving performance had previously been identified; that is, a more dependent relationship in companion dogs fostered impaired problem-solving performance compared with working dogs (Topal et al. 1997).

Yet given the social cognitive abilities of the domestic dog, it is possible that even highly trained dogs might respond to subtle, unintentional handler cues. Dogs' biases for utilizing human movements or social cues impair decision-making and reasoning abilities (Erdohegyi et al. 2007). Dog behavior is further affected by owner/handler gender and personality (Kotrschal et al. 2009). Moreover, dogs evaluate attentional cues of their owners through cues including eye contact and human eye, head and body orientation (Schwab and Huber 2006). Dogs can further distinguish the focus of human attention, using other visual cues such as pointing, gazing, head nodding in the direction of a target, glancing at a target and head turns toward a target affect selection of a target object by a dog (Soproni et al. 2001; Viranyi et al. 2004). In fact, nonverbal cues including proximity of the human to the dog and contextual learning of verbal commands have been shown to moderate dog response to verbal commands (Fukuzawa et al. 2005).

For scent detection dog handlers, beliefs that scent is present might result in either sufficient inadvertent postural and facial cues so that dogs will respond regardless of the absence of scent, beliefs that dogs are providing their trained alert response or simply beliefs that alerts should be called regardless of dog behavior. All of these effects would result in false alerts identified by handlers. These handler beliefs might be influenced by human communication regarding target scent location. Alternatively, handler beliefs might be influenced by increased dog interest in a nontarget scent. The main questions of this study were to (1) determine whether handler beliefs affect detection dog outcomes and (2) evaluate relative importance of dog versus human influences on those beliefs. The present study attempted to determine whether handler beliefs of target scent location would affect outcomes in scent detection dog searches. Importantly, this study was not evaluating abilities of these detection dogs to detect their target scents. Because all dogs were certified, many with confirmed deployment finds their ability to correctly locate target scent was considered to be previously established. Therefore, in order to evaluate outcomes solely based on handler beliefs and expectations, this study was designed so that any alert issued would be a "false" alert; that is, there was no target scent present in any searches conducted for the purposes of this study.

Materials and methods

Handler/dog teams

A total of 18 handler/detection dog teams, recruited through word-of-mouth from multiple agencies, participated in this study. These teams were certified by a law enforcement agency for either drug detection ($n = 13$), explosives detection ($n = 3$), or both drug and explosives detection ($n = 2$). Demographic details of teams, including dog age, dog breed, dog years of detection experience and handler years of detection experience are presented in Table 1. Upon detection of target scent, all explosives dogs, both drug/explosives dogs and one drug detection dog were trained to issue a passive alert; that is, the dog would sit at the location of target scent detection. One drug detection dog was trained to issue a passive–active alert (sitting and barking), and all remaining drug dogs were trained to issue an active alert (barking) upon detection of target scent. All drug detection teams and two teams trained to find explosives had successfully identified their target scents in law enforcement deployment situations. Additional demographic information collected included handler years of experience handling detection dogs, dog years of scent detection experience, dog age and handler-reported breed of dog. In order to maintain

Table 1 Demographic data, n = 18 dog/handler teams

	Day	1	2	All
Dog sex	Male intact	4	9	13
	Male neutered	1	0	1
	Female intact	2	1	3
	Female spayed	0	1	1
Dog breed	GSD	2	1	3
	Labrador	1	0	1
	Belgian malinois	3	5	8
	Dutch shepherd	0	2	2
	Mix	1	3	4
Dog age (years)	Mean	5.0	7.2	6.4
	Median	4.0	6.0	5.8
	Low	2.0	5.0	2.0
	High	10.0	11.0	11.0
Handler scent experience (years)	Mean	5.6	4.0	4.6
	Median	2.0	3.0	3.0
	Low	1.0	1.0	1.0
	High	18.0	7.0	18.0
Dog scent experience (years)	Mean	2.2	3.3	2.9
	Median	1.3	2.0	1.5
	Low	1.0	0.4	0.4
	High	5.0	7.0	7.0

confidentiality, and so that individual teams could not be identified through demographic information, these data were collected anonymously and cannot be linked to any performance data. Due to subject availability, this study was completed across 2 days, with seven teams completing the experiment on the first day, and the remaining 11 teams completing the experiment on the second day.

Procedures

The experimental paradigm in this study was based on a paradigm previously applied to evaluate response conflict in disaster search dogs (Lit and Crawford 2006). Handlers conduct a series of short searches for their target scent across different search scenarios, each representing a different experimental condition. In the current study, there was no target scent present, so that any alert identified by handlers was considered a false alert.

Handler beliefs were influenced either by verbally communicating to the handlers that a specific marker was an indicator of scent location (i.e., human influence), by encouraging dogs to display unusual interest in a specific location with a decoy scent (i.e., dog influence), or by a specific marker that actually indicated the location of a decoy scent (combined human and dog influence). A 4-way single factor experimental design was used to test effects of these influences on handler beliefs. The independent variable was search condition, a within-subjects variable with four levels:

1. *NULL* Unmodified.
2. *MARKED NULL* A piece of 8–1/2" × 11" red construction paper was taped to the door of a cabinet.
3. *UNMARKED DECOY* Two Slim-Jim sausages (removed from their wrappers and stored with their wrappers in an unsealed plastic bag) and a new tennis ball were hidden in the bottom of a pot and placed in a metal cabinet with the doors closed.
4. *MARKED DECOY* Two Slim-Jim sausages (removed from their wrappers and stored with their wrappers in an unsealed plastic bag) and a new tennis ball were hidden in a covered metal electric fryer, which was marked with a piece of red construction paper taped to the outside of the fryer. To minimize the possibility that decoy scents in UNMARKED DECOY and MARKED DECOY were not equally detectable and to encourage dog interest in the decoy scents, the sausages were rubbed along the outside of the cabinet (UNMARKED DECOY) and the electric fryer (MARKED DECOY).

Search conditions were four rooms within a church that had not previously been used for detection dog training purposes. Each room was approximately 30–40 m^2 and contained cabinets, tables and chairs and art supplies. Each condition was identified only as A, B, C or D, indicated by a paper taped on the outside of the door of each room. The experimenter did not touch any items around the rooms, except to place the decoy scents and/or paper markers. To avoid contamination of paper markers with decoy scents, paper markers were placed prior to placement of decoy scents. In order to maintain the belief that the experimenter was setting out target scents in each condition, at the beginning of each testing day, the experimenter carried a metal box containing 12 half-ounce samples of marijuana triple bagged in sealed plastic bags, and a canvas bag containing 12 half-ounce samples of gunpowder triple bagged in sealed plastic bags. Upon entering each condition, the experimenter immediately set these containers down by the door. The experimenter did not handle the scents, and the containers were never opened inside the church. Decoy scents and paper markers were never in contact with these containers and were kept in a separate briefcase carried by the experimenter.

Dog/handler teams completed two searches (maximum 5 min each) in each of the four search areas, for a total of eight trials ("runs") per team. Handlers were provided with a small card containing their assigned sequences of their eight runs, randomly counterbalanced across participants and search areas. Additional written and verbal instructions were provided to handlers that each condition might contain up to three target scents and that target scent markers consisting of a red piece of construction paper would be present in two conditions. No information was provided about the decoy scent.

Each condition had a single observer present. Prior to each search, handlers would indicate to the observer whether their dog was a drug or explosives dog and whether their dog issued a passive or active alert. When a handler "called an alert," that is, confirmed that the dog had found a target scent location and was issuing its trained operant response, the observer would record time of alert and alert location specified by the handler. In marked conditions, if handlers called alerts on the location marked by the paper, observers would record an M to reflect this. Observers recorded alerts as called by handlers and did not evaluate validity of alerts. The same rooms were used for both days of testing. Decoy scents and markers were removed at the end of the first day of testing, and identical but previously unused decoy scents and markers were used for the second day of testing.

This study was double-blind. Neither handler/dog teams nor observers were aware of the conditions of each search area. Because the study was completed across 2 days and we did not want to jeopardize the double-blind nature of this study, all handlers were debriefed and told about the contents of each condition upon the completion of the second day of testing. The experimenter (L. Lit) was the only person present who was aware of the conditions of each search area.

Dependent variables were total number of alerts issued by each dog as reported by handlers in each search area. The correct score for each search area was 0. All alerts were false alerts.

The Institutional Review Board and Animal Care and Use Committee at the University of California at Davis approved this study, and all participants provided written consent.

Statistical analyses

Data were analyzed using SPSS Version 17.0.1. All analyses used a significance threshold of $\alpha < 0.05$ (two-tailed). An omnibus mixed ANOVA was conducted to evaluate effects of day of testing (between groups) and condition (repeated measures) on number of alerts. To evaluate effects of handler influence and dog influence, data were also analyzed as a repeated measures 2 × 2 ANOVA [handler influence (yes/no) and dog influence (yes/no)]. Paired t tests were used to compare alerts between first and second runs of each condition. A chi-squared goodness of fit test compared clean runs (runs with no alerts) in unmarked and marked conditions. Within the MARKED NULL, UNMARKED DECOY and MARKED DECOY conditions, a log likelihood analysis was used to compare runs for which (1) alerts included either a marker or the unmarked decoy scent, (2) alerts did not include the marker or unmarked decoy scent and (3) no alerts were issued, followed by chi-squared goodness of fit tests to compare distribution of these within conditions.

Results

In order to evaluate effects of handler beliefs and expectation on detection dog performance, this study measured performance of 18 handler/dog teams in four separate search areas (NULL, MARKED NULL, UNMARKED DECOY, MARKED DECOY, described in "Materials and methods"). Each team ran each search area twice, for a total of 36 runs per condition (2 runs/team × 18 teams) and an overall total of 144 separate runs (4 search areas × 2 runs/team/area × 18 teams) (Fig. 1).

Day of testing and condition group differences

Overall, because multiple alerts per team within a condition were possible, there were a total of 225 alerts issued. There were 21 (15%) clean runs and 123 (85%) runs with one or more alerts. The omnibus mixed ANOVA using the model "number of alerts = day of testing (between groups) + condition (within-subjects) + [day of testing * condition]" revealed no difference in mean alerts between teams running on the first and second days, $F(1, 16) = 0.94$, $P = 0.35$; no

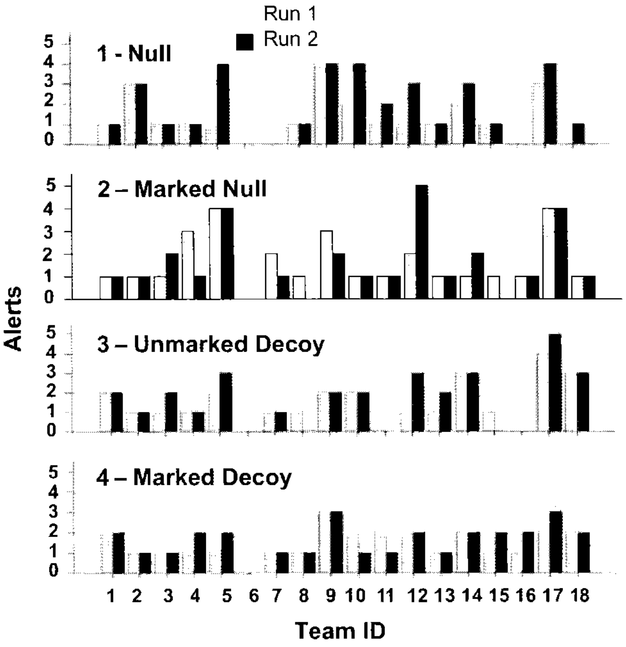

Fig. 1 Alerts for each team across each condition for Run 1 (*light bars*; n = 18/condition) and Run 2 (*dark bars*; n = 18/condition)

difference in mean alerts across conditions, $F(3,48) = 0.09$, $P = 0.97$; and no interaction, $F(3, 48) = 0.63$, $P = 0.60$. Data from both days were subsequently combined for further analysis. The repeated measures 2 × 2 factorial ANOVA found no main effect of human influence, $F(1, 17) = 0.06$, $P = 0.81$; no main effect of dog influence, $F(1, 17) = 0.01$, $P = 0.93$; and no interactions between human influence and dog influence, $F(1, 17) = 0.01$, $P = 0.94$.

First and second run differences

Within each condition, there was no difference in mean alerts between the first and second runs, except for NULL, where there were more alerts on the second run compared with the first run (paired $t[17] = -2.83$, $P = 0.01$).

Effect of marker on clean runs

Distribution of clean runs differed across unmarked and marked areas. There were more clean runs in unmarked areas (NULL and UNMARKED DECOY combined) ($n = 15$) than in marked areas (MARKED NULL and MARKED DECOY combined) ($n = 6$), $X^2[1, 21] = 3.86$, $P = 0.05$. In contrast, distribution of clean runs was not different across runs with and without decoy scent (NULL and MARKED NULL combined, $n = 11$, compared with UNMARKED DECOY and MARKED DECOY combined, $n = 10$), $X^2[1, 21] = 0.05$, $P = 0.827$.

Human and dog influences on alert locations

Alert locations in conditions marked with paper (MARKED NULL), containing decoy scent (UNMARKED DECOY) and containing decoy scent marked with paper (MARKED DECOY) were compared to evaluate differences of human influence on handler beliefs and dog influence on handler beliefs. Runs were grouped according to whether any one of the alerts in that run (1) included the marker and/or decoy scent; (2) did not include the marker and/or decoy scent; or (3) the run was clean (no alerts). These groups were dependent on condition, log likelihood $[4, 108] = 22.236$, $P < 0.001$, $\Phi = 0.41$ (Fig. 2). There were significantly more runs including alerts on the marker than either clean runs or runs not including alerts on the marker in both MARKED NULL ($X^2[1, 36] = 21.78$, $P < 0.001$) and MARKED DECOY ($X^2[2, 36] = 36.5$, $P < 0.001$) (Fig. 2). This was different than UNMARKED DECOY, where there were no differences between clean runs, runs with alerts on the decoy scent and runs not including alerts on the decoy scent ($X^2[2, 36] = 4.67$, $P = 0.09$) (Fig. 2). Conversely, comparing across conditions (black bars, Fig. 2), there were more runs with alerts on marked locations in MARKED NULL and MARKED DECOY than UNMARKED DECOY, although the differences were not significant when corrected for multiple comparisons (Fig. 2).

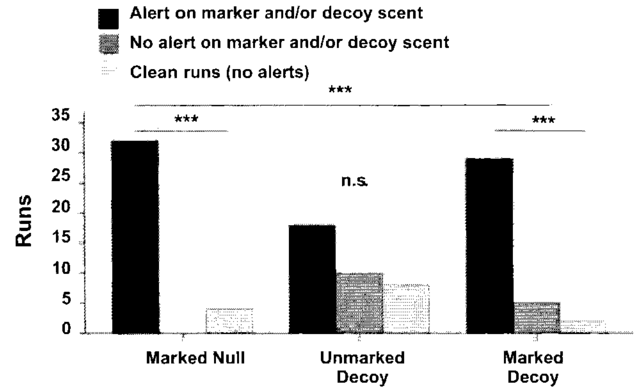

Fig. 2 Runs within each condition (combined $n = 36$) with alerts including marker and/or decoy scent (*black bars*), not including marker and/or decoy scent (*dark gray bars*), or clean runs (*light gray bars*). Asterisks represent statistically significant differences between groups as shown by log likelihood (across all conditions) and chi-squared test (within conditions); $^{***}P < 0.001$; n.s. not significant

Trend analysis

Finally, counterbalancing run order across participants ensured that each participant ran conditions in a different order. To evaluate whether there was an effect of sequence order of runs on alerts, all runs were reordered to reflect the sequence in which participants completed the conditions. Trend analysis was performed relating condition order to the number of alerts per run. An analysis of the cubic component of trend was significant, $F(1, 17) = 7.67$, $P = 0.01$, $\eta_p^2 = 0.31$, indicating that this trend accounted for over one-third of the variance in number of alerts per run (Fig. 3, solid line). This trend was consistent across both days of testing (Fig. 3, dotted and dashed lines).

Discussion

The goals of this study were to (1) identify whether handler beliefs affect detection handler/dog team performance and (2) evaluate relative importance of dog versus human inputs on those beliefs. To test this, we influenced handler beliefs and evaluated subsequent handler/dog team performance according to handler-identified alerts. The overwhelming number of incorrect alerts identified across conditions confirms that handler beliefs affect performance. Further, the directed pattern of alerts in conditions containing a marker compared with the pattern of alerts in the condition with unmarked decoy scent suggests that human influence on handler beliefs affects alerts to a greater degree than dog influence on handler beliefs. That

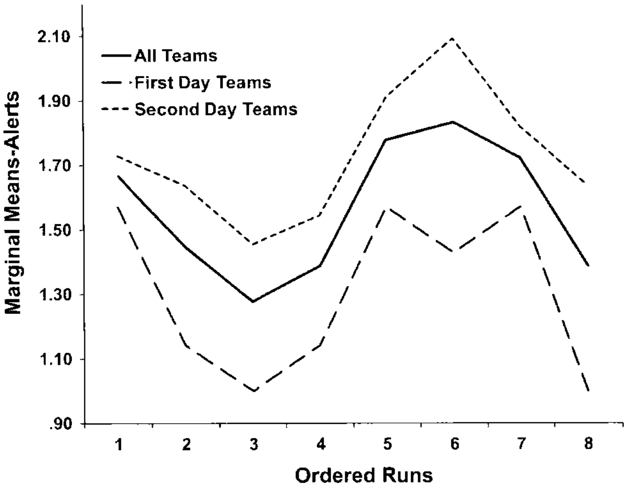

Fig. 3 Cubic trend for all teams (*solid black line*, $n = 18$) relating condition run order (ordered runs) to marginal means of alerts per run as shown by trend analysis, $P = 0.01$, $\eta_p^2 = 0.31$. Trends for teams from first day (*dashed line*, $n = 7$) and second day (*dotted line*, $n = 11$) are also shown for comparative purposes

is, total number of alerts identified by handlers did not differ across conditions. However, distribution of these alerts did differ across conditions; more alerts were identified on target locations indicated by human suggestion (paper marker) than on locations indicated by increased dog interest (hidden sausage and tennis balls).

In light of written and verbalized instructions that "Each scenario may contain up to 3 of your target scents," it was interesting that there were 12 runs with either four or five alerts (Fig. 1). It was unclear whether handlers did not attend to the instructions, did not remember the instructions or believed that there were more than three target scent sources in each condition.

There are two possible explanations for the large number of false alerts identified by handlers. Either (1) handlers were erroneously calling alerts on locations at which they believed target scent was located or (2) handler belief that scent was present affected their dogs' alerting behavior so that dogs were alerting at locations indicated by handlers (that is, the Clever Hans effect).

In the event that handlers were indeed asserting dog alerts regardless of dog response (or lack thereof), there are two possible causes. The handlers' beliefs that scent was present may have been sufficient motivation to identify alerts even when the handlers were clearly aware that the dog had not provided the trained alert response behavior. Alternatively, the handlers' beliefs were sufficient to generate a form of confabulation. Broadly defined, confabulation refers to false beliefs that may be unrelated to actual experienced events (Bortolotti and Cox 2009). Information regarding prevalent events (events that are common and therefore of increased likelihood) makes events more self-relevant and increases beliefs in occurrence of such events (van Golde et al. 2010). Thus, the perceived likelihood that scent was present across conditions would have contributed to confidence in handler beliefs of scent and dog responses. Because other-generated suggestions influence beliefs and subsequent actions more strongly than self-generated suggestions (Pezdek et al. 2009), the experimenter-provided suggestion that target scent was present may have further contributed to this effect. However, the conclusion that handlers are asserting their dogs have alerted simply upon seeing the marked areas regardless of actual dog response does not account for the numerous additional alerts occurring in other areas. In addition, the experimenter was informed that three handlers admitted to overtly cueing their dogs to alert at the marked locations, suggesting that handlers would not call alerts unless and until they observe the dogs' trained responses. Handlers are trained to recognize and reward specific behaviors of their dogs. The exhibition of an alert is an obvious and discrete behavior. Although data describing observer assessments were not collected, all observers were familiar with detection dog training and performance, and all observers were visibly surprised upon debrief (L. Lit, personal communication). Therefore, it is unlikely, although cannot be absolutely confirmed, that handlers called alerts on markers without seeing an appropriate behavior from the dog.

It may be more parsimonious to suggest that dogs respond not only to scent, but to additional cues issued by handlers as well. This is especially plausible since, in training, alerts are originally elicited through overt handler cueing. Cueing in initial training may include overt cues, verbal commands and physical prompting. Cues may also include more subtle unintentional cues given by handlers such as differences in handler proximity to the dog according to scent location, gaze and gesture cues, and postural cues.

Human cues that direct dog responses without formal training include pointing, nodding, head turning and gazing (reviewed in Reid 2009). While formal obedience training can enhance dogs' use of human cues (McKinley and Sambrook 2000), type of training can differentially affect dogs' human-directed communicative behaviors (Marshall-Pescini et al. 2009, 2008). Gazit et al. (2005) found diminished response when an area searched repeatedly was lacking target scent. While the proposed reason for their findings emphasized effects of context specificity on the detection dogs (Gazit et al. 2005), the current findings raise the possibility that at least some of the effects of Gazit et al. (2005) might have arisen due to handler beliefs that scent would not be present in that area, with subsequent attenuation of dog response.

Because the current study did not include videotape of handler/dog team performance, there is no way to identify

Table 2 Alert locations and alert frequencies (#) in each location for all scenarios

NULL		MARKED NULL		UNMARKED DECOY		MARKED DECOY	
Alert location	#	Alert location	#	Alert location	#	Alert location	#
Air conditioner	11	MARKER	32	DECOY SCENT	18	MARKER	29
First-aid kit	10	Easel	9	Piano	15	Clear bin	12
Wall heater	9	Tall cabinet	6	Wall heater	7	Oven	3
Right window	7	Cart	3	Red bag	6	Tool box	3
Tall cabinet	5	Chalkboard	3	Radiator	5	Gray tote	2
Desk	4	Blinds	1	Upholstered chair	3	Above boxes	1
Short cabinet	4	Desk chair	1	Shelf	1	Back table	1
Trash can	4	Pedestal	1	Table	1	Doorway	1
Map on chalkboard	1	Trash can	1			Painted box	1
Pencil sharpener	1					Paint container	1
Table	1					Trash can	1
Totals	57		57		56		55

which conclusion would be appropriate. Observer coding of dog behavior was not likely to improve the reliability of the data acquired because the double-blind study design had the potential for the observers to be subject to the same biases as the handlers. In fact, it is possible that the observers were subject to greater biases than the handlers, since they were able to observe every dog twice. Therefore, observer coding would have been subject to the same possible explanations as the handlers, and further subject to question according to level of observer experience with working dogs. Future studies should directly explore underlying factors responsible for the false alerts as this will improve development of effective remedies to optimize performance.

Dogs can learn to respond to human gestures very rapidly (Bentosela et al. 2008; Elgier et al. 2009; Udell et al. 2008). Thus, it is tempting to speculate that the large number of false alerts resulted from reinforcement of dogs for false alerts received in earlier conditions. However, the pattern of alerts, consistent across days of testing (Fig. 3), suggests that alerts did not reflect a simple learning effect. This is supported by prior studies of human–dog social cognitive interactions demonstrating no clear learning effect when comparing early with later trials (Hare et al. 2002; Riedel et al. 2008).

When considering alternative explanations for the incorrect responses, it is further possible that some alerts resulted from target scent contamination during initial setup of conditions. This is unlikely, given the emphasis of alerts toward marked sites, particularly when considering that the pattern of alerts was modified by human influence. The array of alert locations (Table 2) also does not support this explanation, notably because no dogs alerted on or around the doors where the scent containers had briefly been placed. Moreover, detection dogs are trained to identify scent source rather than scattered residual scent. For example, dogs trained to alert on gunpowder are not expected to alert in an airport area simply because an armed officer passes through. The significant trend (Fig. 3) further suggests that a temporal component contributed to the number of alerts under these experiments.

It is possible, although also unlikely, that all objects in the room smelled like the dogs' target scents. Because these were rooms in a church building that had not previously been used for detection dog training, it was also unlikely that there were explosives or drugs that had been stored within the testing rooms. Some handlers suggested the possibility that dogs were following previous dogs and alerting at locations in which these dogs had salivated or otherwise left trace evidence of their presence. This would not explain the difference in patterns of alerts between marked and unmarked conditions or the variation in alert locations across all conditions. This would also be unlikely given the extensive training and certification processes required of these teams.

It is important to emphasize that this study did not evaluate performance of dogs when presented with scent. Handler-dog teams undergo substantial training and rigorous certification prior to deployment; all teams included in this study confirmed prior successful finds during active deployment. This study only considered number of alerts under the artificially manipulated condition of handler belief of scent when in fact no scent was present.

In conclusion, these findings confirm that handler beliefs affect working dog outcomes, and human indication of scent location affects distribution of alerts more than dog interest in a particular location. These findings emphasize the importance of understanding both human and human–dog social cognitive factors in applied situations.

Acknowledgments We would like to acknowledge the handlers and their dogs who agreed to participate in this research and observers who helped in the data collection process. This work was supported by an Autism Researcher Training Program fellowship (LL, T32 MH073124) and a private contribution.

References

Bentosela M, Barrera G, Jakovcevic A, Elgier AM, Mustaca AE (2008) Effect of reinforcement, reinforcer omission and extinction on a communicative response in domestic dogs (*Canis familiaris*). Behav Processes 78:464–469

Bortolotti L, Cox RE (2009) 'Faultless' ignorance: strengths and limitations of epistemic definitions of confabulation. Conscious Cogn 18:952–965

Elgier AM, Jakovcevic A, Barrera G, Mustaca AE, Bentosela M (2009) Communication between domestic dogs (*Canis familiaris*) and humans: dogs are good learners. Behav Processes 81:402–408

Erdohegyi Á, Topál J, Virányi Z, Miklósi Á (2007) Dog-logic: inferential reasoning in a two-way choice task and its restricted use. Anim Behav 74:725–737

Fukuzawa M, Mills DS, Cooper JJ (2005) More than just a word: non-semantic command variables affect obedience in the domestic dog (*Canis familiaris*). Appl Anim Behav Sci 91:129–141

Gazit I, Goldblatt A, Terkel J (2005) The role of context specificity in learning: the effects of training context on explosives detection in dogs. Anim Cogn 8:143–150

Hare B, Brown M, Williamson C, Tomasello M (2002) The domestication of social cognition in dogs. Science 298:1634–1636

Horvath G, Jarverud GA, Jarverud S, Horvath I (2008) Human ovarian carcinomas detected by specific odor. Integr Cancer Ther 7:76–80

Kotrschal K, Schoberl I, Bauer B, Thibeaut AM, Wedl M (2009) Dyadic relationships and operational performance of male and female owners and their male dogs. Behav Processes 81:383–391

Lit L (2009) Evaluating learning tasks commonly applied in detection dog training. In: Helton W (ed) Canine ergonomics: the science of working dogs. CRC Press, Boca Raton

Lit L, Crawford CA (2006) Effects of training paradigms on search dog performance. Appl Anim Behav Sci 98:277–292

Marshall-Pescini S, Valsecchi P, Petak I, Accorsi PA, Previde EP (2008) Does training make you smarter? The effects of training on dogs' performance (*Canis familiaris*) in a problem solving task. Behav Processes 78:449–454

Marshall-Pescini S, Passalacqua C, Barnard S, Valsecchi P, Prato-Previde E (2009) Agility and search and rescue training differently affects pet dogs' behaviour in socio-cognitive tasks. Behav Processes 81:416–422

McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T (2006) Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. Integr Cancer Ther 5:30–39

McKinley J, Sambrook TD (2000) Use of human-given cues by domestic dogs (*Canis familiaris*) and horses (*Equus caballus*). Anim Cogn 3:13–22

Miklosi A, Topal J, Csanyi V (2007) Big thoughts in small brains? Dogs as a model for understanding human social cognition. Neuroreport 18:467–471

Oesterhelweg L, Krober S, Rottmann K, Willhoft J, Braun C, Thies N, Puschel K, Silkenath J, Gehl A (2008) Cadaver dogs–a study on detection of contaminated carpet squares. Forensic Sci Int 174:35–39

Pezdek K, Lam ST, Sperry K (2009) Forced confabulation more strongly influences event memory if suggestions are other-generated than self-generated. Leg Criminol Psychol 14:241–252

Pfungst O (1911) Clever Hans (The horse of Mr. von Osten): a contribution to experimental animal and human psychology (C. L. Rahn, Trans.). Henry Holt, New York

Prato-Previde E, Marshall-Pescini S, Valsecchi P (2008) Is your choice my choice? The owners' effect on pet dogs' (*Canis lupus familiaris*) performance in a food choice task. Anim Cogn 11:167–174

Reid PJ (2009) Adapting to the human world: dogs' responsiveness to our social cues. Behav Processes 80:325–333

Riedel J, Schumann K, Kaminski J, Call J, Tomasello M (2008) The early ontogeny of human-dog communication. Anim Behav 75:1003–1014

Schwab C, Huber L (2006) Obey or not obey? Dogs (*Canis familiaris*) behave differently in response to attentional states of their owners. J Comp Psychol 120:169–175

Soproni K, Miklosi A, Topal J, Csanyi V (2001) Comprehension of human communicative signs in pet dogs (*Canis familiaris*). J Comp Psychol 115:122–126

Szetei V, Miklósi Á, Topál J, Csányi V (2003) When dogs seem to lose their nose: an investigation on the use of visual and olfactory cues in communicative context between dog and owner. Appl Anim Behav Sci 83:141–152

Topal J, Miklosi A, Csanyi V (1997) Dog-human relationship affects problem solving behavior in the dog. Anthrozoos 10:214–224

Udell MAR, Giglio RF, Wynne CDL (2008) Domestic dogs (*Canis familiaris*) use human gestures but not nonhuman tokens to find hidden food. J Comp Psychol 122:84–93

van Golde C, Sharman SJ, Candel I (2010) High prevalence information from different sources affects the development of false beliefs. Appl Cogn Psychol 24:152–163

Viranyi Z, Topal J, Gacsi M, Miklosi A, Csanyi V (2004) Dogs respond appropriately to cues of humans' attentional focus. Behav Processes 66:161–172

Wasser SK, Davenport B, Ramage ER, Hunt KE, Parker M, Clarke C, Stenhouse G (2004) Scat detection dogs in wildlife research and management: application to grizzly and black bears in the Yellowhead Ecosystem, Alberta, Canada. Can J Zool-Revue Canadienne De Zoologie 82:475–492

APPENDIX I

THE CHAMPION National Association of Criminal Defense Lawyers

Champion Magazine

NATIONAL ASSOCIATION OF CRIMINAL DEFENSE LAWYERS

Who We Are
Press Room
News & Issues
CLE & Events
Committees
Members Only
Champion Magazine
Indigent Defense
Federal Legislation
White Collar
State Legislation
Affiliate Organizations
Lawyer Resources
Foundation
NACDL.org

Print or Email This Page

April 2006, Page 12

Those Doggone Sniffs Are Often Wrong: The Fourth Amendment Has Gone To The Dogs

By Jeffrey S. Weiner; Kimberly Homan

Perpetuating the myth of the precision and noninvasiveness of the "*sui generis*" dog sniff which informed the Supreme Court's decision in *United States v. Place*,[1] in which the Court held that the brief detention of luggage located in a public place for purposes of exposing it to a dog sniff did not constitute a search within the meaning of the Fourth Amendment,[2] the Court in *Illinois v. Caballes*[3] held that the employment of a dog sniff during a legitimate traffic stop does not constitute a constitutionally cognizable intrusion upon legitimate privacy interests.[4] The purpose of this article is not to critique *Caballes*' myopic Fourth Amendment analysis, which has already been ably done in this publication and elsewhere,[5] nor is it to explore the current state of the law regarding the legal significance of so-called positive canine "alerts" for reasonable suspicion, probable cause, stops, seizures, or searches. Instead, it has a more practical focus, one which concentrates on the litigation of motions to suppress in which the issue is whether a canine "alert" was sufficiently reliable to establish probable cause for the search. This article will explore various ways in which, through litigating motions to suppress the fruits of searches conducted as the result of a positive dog "alert," defense counsel can educate courts, which are all too apt to reflexively credit the accuracy of canine alerts, regarding the serious flaws in much of the existing jurisprudence regarding canine sniffs and regarding modes of inquiry which more accurately reflect the reality of the interaction between dog, handler, and target. Given the entrenched myth of the infallible canine, this will be a long, frequently thankless, process, often bringing little satisfaction in individual cases, but it is one which must be pursued if rationality is to assert itself in this area of the law. Just as repeated litigation has exposed the flaws of other firmly-accepted law enforcement techniques, so, too, may the canine alert come to be regarded in its proper perspective.

For present purposes, *Caballes* is most noteworthy not for its result, which was, unfortunately, all too predictable, but for Justice Souter's dissenting opinion, in which he sought to dispel the myth of the infallible canine, a crucial underpinning of the Court's opinion, which he rightly labeled "a creature of legal fiction:"[6]

At the heart of both *Place* and the Court's opinion today is the proposition that sniffs by a trained dog are *sui generis* because a reaction by the dog in going alert is a response to nothing but the presence of contraband. . . . Hence, the argument goes, because the sniff can only reveal the presence of items devoid of any legal use, the "sniff does not implicate legitimate privacy interests" and is not to be treated as a search. . . .

> The infallible dog, however, is a creature of legal fiction. Although the Supreme Court of Illinois did not get into the sniffing averages of drug dogs, their supposed infallibility is belied by judicial opinions describing well-trained animals sniffing and alerting with less than perfect accuracy, whether owing to error by their handlers, the limitations of the dogs themselves, or even the pervasive contamination of currency by cocaine. . . .

> Indeed, a study cited by Illinois in this case for the proposition that dog sniffs are "generally reliable" shows that dogs in artificial testing situations return false positives anywhere from 12.5 to 60 percent of the time, depending on the length of the search. . . . In practical terms, the evidence is clear that the dog that alerts hundreds of times will be wrong dozens of times.7

Positive canine alerts are treated as virtually — if not completely — *per se* probable cause in virtually every circuit, at least if the canine is shown to be "trained" and "reliable".8 While those words appear with monotonous and unthinking regularity in judicial opinions relating to dog sniffs, handed down from opinion to opinion over the years, courts have, with a relative paucity of exceptions, demonstrated only an, at best, incomplete understanding of what these concepts actually mean in the real world and an entrenched disinclination to look much beyond the facts that the dog's handler testified to the occurrence of an alert and that the dog was trained and certified.

There are two primary deficiencies in this narrow focus.

First, it largely, if not entirely, ignores the role of the dog's handler, generally a law enforcement officer and typically a local police officer or sheriff.

Second, and relatedly, it assumes that an "alert" is an alert because the handler said it was. Most courts have failed to consider — or even recognize — the role of the dog's handler in the process. The handler is not simply someone who holds the leash while the dog walks around and sniffs. Instead,the dog and handler function as an integral team. The dog is the sensor, and the handler is the trainer and interpreter. The handler's performance in both roles is inseparably intertwined with the dog's overall reliability rate. . . . And since the net result is the product of the interaction between two living beings, both roles of the handler are highly subjective.9

Thus, the proper focus of the reliability inquiry is not simply upon the training and certification of the dog,10 but on the training and reliability of the dog/handler team.

Role Of The Handler

Dogs can be trained to distinguish various odors with some degree of accuracy, but they are far from infallible, and their responses are often only as reliable as the interpretations placed upon them by their handlers.11 In fact, one commentator has reported that on most occasions on which no drugs are found as the result of a search based on a canine alert, the fault is not that of the dog but rather that of the handler in misinterpreting the dog's behavior.12 While a dog can be initially trained in a few weeks, training the human handler requires more time and effort.13 In some circuits, an affidavit in support of an application for a search warrant based in whole or in part upon a positive canine alert need do no more to establish the dog's reliability than to state that the dog is "trained" and/or "certified."14 This is, however, only half the equation. The focus must be redirected to the reliability and integrity of the dog/handler *team*. The amount of training the handler has had and the length of time the handler and dog have worked together are important factors, and the reported cases indicate that handler training runs the gamut from two weeks to hundreds of hours.15

The amount of training and experience which the handler has had with the dog is a critical element, as, contrary to the way in which courts have tended to view the process, the dog does not announce to the handler that, if he searches, he will find drugs in a particular location. Rather, each dog behaves in a particular manner, which the handler then "interprets" as an "alert," indicating the presence of drugs. An alert is not the objective proof as which it has generally been regarded by the courts; instead, it has a large — and sometime *very* large — subjective component. Some courts have expressed an appreciation of the role which handler interpretation plays in the process, although, with rare exception, that appreciation has not led to any less deference to the handler's testimony that the dog alerted.16 Repeated stress upon the subjective and interpretive aspects of the canine inspection, bolstered by expert testimony should dislodge the unshakable faith which courts have placed in the objective veracity of canine alerts.

Alert

Another too-little examined aspect of alert-based searches is the question of just what is an "alert." Dogs are trained to give a particular response when they smell certain illegal drugs and, during training, are rewarded when they correctly give that response in the presence of drug odor.17 In general, dogs are trained to alert "aggressively" or "passively," responses which have very different physical manifestations.18 When the handler testifies that the dog alerted in something other than the manner in which it was trained to respond to the presence of drugs, there is cause to question whether the dog actually did alert, *i.e.*, give the final response which it was trained to make when it was certain that the odor it had been trained to detect was present and, concomitantly, reason to decline to base a finding of probable cause to search on the alert alone. As Dr. Daniel Craig, a noted expert in canine training and performance, who has on a number of occasions testified as a defense expert witness in dog-sniff cases, has explained:

> Detector dog handlers have been known to say things like "I can read my dog," "My dog knows it's there," "My dog's behavior tells me it's in there," "I can read my dog's behavioral change and I know the odor is there," "I am the only one who can read my dog," "I know what my dog is thinking," "I know when he is in the scent cone," et cetera. Are they just repeating what they were taught? If not, where do they get this notion? In initial training and subsequent training the only time they reward (reinforce) their dog is when the dog makes the definitive defined final response. Then and only then can the trainer verify that the dog has detected and responded to a specific target odor. The dog is rewarded for that response and no other.
>
> The first thing one must do . . . is decide what specific response the dog must make in order to determine if it is responding correctly to a selected target odor. . . . The handler or trainer must be able to articulate that specific response to anyone not in the dog training profession. That specific response is the only response you reward with the selected primary reinforcement. . . .
>
> If the dog does not make the defined final response sometime during a search, the target odor is either not present or the dog or handler made an error. Dogs do respond when no target odor is present. They also fail to respond when a target odor is present. The handler may assume any response other than the defined final response verifies the presence of the target odor. At this point the handler is guilty of interpretation, supposition, or speculation. The dog has the olfactory sensing system (nose) and the final decision as to the presence or absence of a target odor is up to the dog and not the handler. A well-trained detector dog will only respond to the target odor(s) it has been properly trained to detect. That dog will emit the defined final response it was trained to make to a target odor at a predetermined rate of accuracy.
>
> Educated guesses based upon the handler's knowledge of their dog's training and past performance are nothing more than educated guesses when their dog fails to make the defined final response during a specific search. . . .[19]

Thus, it is not enough that the handler testifies at the suppression hearing that the dog "alerted," and defense counsel should hesitate before stipulating that the dog alerted and should refrain from conceding that an "alert" is a reliable indicator of anything. Instead, the handler should be examined at the suppression hearing or, where permitted, in pre-hearing deposition, to elicit a precise description of the definitive final response which the dog was trained to give upon identifying the odor of narcotics. If the dog gave any other response, then it was not doing what it was trained to do, and there is sound reason to question whether an alert did in fact take place.[20] There may be videotapes of the dog's training exercises available to assist in this process,[21] just as there may be a videotape of the stop during which the canine sniff was conducted,[22] both of which may provide crucial evidence, at least if the "alert" appears on the videotape, which — unsurprisingly — it sometimes does not.[23] Unfortunately, and incredibly, most courts have blindly credited testimony from handlers of the "I know it when I see it" variety. [24]

A few courts have, however, wisely been skeptical of reliance upon canine behavior which differed from the dog's trained final response and depended entirely upon the handler's interpretation of equivocal behavior. In *United States v. Heir*,[25] for example, the dog was trained to alert by scratching, but had not done so in this case. Instead, the dog sniffed intently around certain areas of the car, which the handler testified constituted an alert, acknowledging, however, that such alert behavior was "subtle" and might only be recognized by himself or another person familiar with the dog's tendencies. Defense experts testified that they saw nothing on the videotape of the stop to indicate that the dog had alerted. The court adopted the magistrate judge's conclusion that an alert had not occurred, as well as the conclusion that even if the alert behavior described by the handler occurred, it was too subjective a standard to establish probable cause. Instead, the court ruled, an "objectively observable 'indication' by the dog of the presence of drugs" was required.[26]

Then there is the problem of "cuing," conduct of the handler during the dog sniff which, consciously or unconsciously, influences the reaction of the dog and may prompt an "alert" reaction from the dog in response to the handler's cues rather than to the presence of drugs. One important aspect of handler training is the avoidance of such cues.[27] While courts have recognized the potentially pernicious effects of cuing, the ability of defendants to establish that cuing occurred has proven elusive. For example, in *United States v. Trayer*,28 the defendant's expert testified that it is possible for the handler to compromise the dog's supposedly infallible objectivity through voice or physical cues and further testified that this sniff — along a train corridor outside the sleeping compartments — did not satisfy the best standards, as the dog was able to observe that the handler went only as far down the corridor as the defendant's compartment before returning to the dog and bringing the dog down the corridor.29

The court found this testimony "quite troubling" in light of the reliance which courts place on canine alerts in probable cause determinations, but, typically, ultimately upheld the search, as the defendant's expert was unwilling to say from the evidence that the handler actually cued the dog.30

The court did, however, express itself "mindful that less than scrupulously neutral procedures, which create at least the possibility of unconscious 'cuing', may well jeopardize the reliability of dog sniffs." 31In *United States v. McLaughlin*, 32the defense expert testified, in essence, that the handler used his hunch that drugs were present to induce the dog to alert. The court agreed that "using a dog search as a sham in finding drugs — essentially not allowing the dog to employ the skill for which he was trained — would significantly decrease the reliability of the search," but, as might be expected, concluded that the record did not support the contention that the dog sniff was simply a ruse. Lacking, the court said, was any evidence that the handler's conduct had an improper effect on the dog. 33In contrast, in *Heir*, in which the dog had responded by sniffing intently but did not give its trained final response, the defendant's experts testified that "the alert behavior described by [the handler] could easily be attributed to his 'cuing' of the animal, either intentionally or unintentionally, by changing the leash from one hand to the other, by stopping, by blocking the way, or by other actions."34

The court was unwilling to accept that an actual alert had occurred. 35Cases like this one are too rare. The Supreme Court should revisit this issue as soon as possible and restore the protections of the Fourth Amendment which have gone to the dogs.

Reliability

"Reliability" is generally regarded by the courts as the measure of how likely an alert by the dog in question is to be an accurate indicator of the presence of drugs (or of whatever the dog has been trained to detect). Dogs are not scientific machines. A number of courts have recognized — or at least paid lip service to — the concept that a particular dog/handler team may be insufficiently reliable to support a finding of probable cause. As one court has stated:

[T]he possibility of error exists and, in limited circumstances, the error may be of such magnitude that a canine alert is not sufficient to establish probable cause. For instance, it stretches the bounds of jurisprudential imagination to believe that a positive alert by an untrained dog or by a dog with an extensive history of false positive alerts could be relied upon to establish probable cause without raising Fourth Amendment concerns.36

In challenging the reliability of the dog/handler team, access to the dog's training, certification, and field performance records is essential. Some courts have recognized the importance of discovery regarding the dog's training, certification, and performance records,37 but others continue to regard exploration of the documentary record of the dog's performance as largely a waste of time, holding that the government need only present the testimony of the handler and need not produce the underlying records.38 The more thoroughly courts can be persuaded to regard alerts by dogs which simply have been "trained" and "certified" as something less than automatic probable cause, the more likely courts will come to regard discovery requests for the historical records of the dog's training and field performance as an essential component of the litigation of motions to suppress where probable cause for the search was predicated in whole or in part on a canine alert.

Access to the records of the dog's performance in training and in the field is essential to challenging predictable handler testimony regarding the always "excellent reliability" of their dogs — as well as to convincing courts that handler testimony alone should not be considered sufficient to establish the reliability of the dog.39 For example, handlers will sometimes report very high — even perfect — accuracy rates, even where drugs have not been found following the dog's alert, based on the unprovable assumption that, if the dog alerted, it must mean that drugs had been present in the location to which the dog alerted.40 Only with the dog's training records can it potentially be determined how often this occurred. Unfortunately, courts have, by and large, been only too ready to accept the premise that "residual odor" provides a valid explanation for a false alert41 or to admit canine alert evidence as evidence that drugs were, at some time, present at the location to which the dog alerted. 42

The training and performance records are also the only way — assuming proper records have been kept (a major assumption) — to ascertain the dog's actual reliability rate, which courts generally regard as the percentage of times the dog's alerts have proven accurate.43 Such records are frequently the only means of establishing an accurate measure of just how often the dog has been right and how often it has been wrong.44

The records may be just as important for what they do not show as for what they do show. For example, it is not enough that a dog has been trained and certified; the dog needs frequent practice to maintain its training,45 and the keeping of careful records is a crucial part of this process.46 As one court has explained:

[The training and certification academy's] manual instructed [the handler] to keep proper records of [the dog's] activities and periodically field train [the dog] to ensure [the dog's] continued reliability. [The academy's] continued assurance of [the dog's] accuracy depended on [the handler] following these instructions. Over time, if not properly monitored, a dog may fall out of its trained behavior and begin responding to a handler's cues rather than to actual detection of a narcotic odor. A drug dog will lose its effectiveness in the field and may revert to old, bad habits if not continually trained. Accurate recordkeeping is essential to ensure the dog's reliability until the dog is recertified.47

Given the fact that in many circuits, an affidavit seeking a search warrant based on a canine alert need say no more than that the dog was trained and/or certified, it is important to stress whenever possible, in cases involving searches with and without warrants, that the fact that the dog *was* trained and certified is no guarantee of continuing reliability. Perhaps, over time, courts will get the message, although why they have not done so yet is unfathomable in light of the manifest unsoundness of equating a handler's testimony that his dog alerted with automatic probable cause.48

Search Warrant Affidavits

So unquestioning is the faith of courts in the accuracy of the "trained" and "certified" dog that in some circuits, the affiant need say no more about the dog to support a finding of probable cause, not even a conclusory statement that the dog is reliable.49 In other circuits, it suffices that the affidavit merely says that the dog has proven reliable in the past.50 Such rulings are entirely inconsistent with the well-established proposition that the judicial officer is to make an *independent* assessment of the existence of probable cause before issuing a warrant and not simply ratify the conclusions of law enforcement officers. Unless the judicial officer is provided with information regarding the dog's performance during training and in the field, he or she cannot make the requisite independent assessment of the likelihood that drugs will be found in the location to which the dog alerted.51 Instead, the probable cause determination will remain solely in the hands of law enforcement officers, where it does not belong.

Even where data regarding the dog's past performance is set forth in the affidavit, affiants frequently mention only the dogs' successes.52 Such data, however, presents an incomplete, and potentially highly misleading, picture. For example, if the affidavit says that actual drugs (not residual odor) were found fifty times when the dog alerted, this would reflect a phenomenal success rate if it had alerted a total of fifty times. The case would be quite otherwise however, if the dog had alerted a hundred times. 53

If the information regarding the dog set forth in the affidavit is inaccurate or materially misleading, or if material information regarding the dog's reliability was omitted from the affidavit, then a challenge under *Franks v. Delaware*54 may be in order. It appears to be generally accepted that *Franks* applies — at least in theory — in the context of canine alert averments.55 A *Franks* challenge will be entertained, however, only if sufficient support for it is demonstrated — another reason why access to the dog's records is so important.

Notes

1. 462 U.S. 696, 707 (1983).
2. *Id.* The *Place* Court did go on to invalidate the seizure of the defendant's luggage as unreasonable, based upon its 90-minute duration while awaiting the arrival of the sniffer dog. *Id.* at 710.
3. 543 U.S. 405, 125 S.Ct. 834 (2005).
4. 125 S.Ct. at 838.
5. *See, e.g.,* M. Hirsch & D.O. Markus, *Fourth Amendment Forum*, 29-JUN Champion 48 (2005).
6. 125 S.Ct. at 839 (Souter, J., dissenting). Some courts have recognized that drug-sniffing dogs are not infallible, although this recognition has not, by and large, had any effect on the ultimate outcome of motions to suppress. *See, e.g.,* United States v. Rosario-Peralta, 199

F.3d 552, 562 (1st Cir. 1999)("Some of the facts elicited by defendants do demonstrate that dog sniff testimony is not a perfect indicator of a particular controlled substance in a particular, well-defined location at a particular time"), *cert. denied sub nom.* Antonio Javier v. United States, 531 U.S. 902 (2000); United States v. Outlaw, 134 F.Supp.2d 807, 813 (W.D.Tex. 2001)(canine inspections not an infallible means of detecting drugs), *aff'd*, 319 F.3d 701 (5th Cir. 2003).

7. *Id.* at 839-40 (citations omitted). As Justice Souter indicated, both *Caballes* and *Place* also rest upon the questionable premise that canine alerts are noninvasive and uniquely calibrated to reveal only the presence of contraband. This proposition is dubious and misguided for two separate reasons. First, the Court ignored the fact that drug-sniffing dogs are large, frequently intimidating, and sometimes quite frightening. *See, e.g., Caballes*, 125 S.Ct. at 845 (Ginsburg, J., dissenting)("A drug-detection dog is an intimidating animal"). Such dogs do, at least at times, invade what most people would regard as a zone of personal privacy. *See, e.g.*, United States v. Kelly, 128 F.Supp.2d 1021, 1024 (S.D.Tex. 2001)(during random canine sweep of persons entering country, dog put nose in defendant's groin), *aff'd* 302 F.3d 291 (5th Cir.), *cert. denied*, 537 U.S. 1094 (2002). Second, while the Court was literally correct that a positive dog alert does not by itself reveal the contents of the vehicle or container sniffed, except insofar as it indicates the possible presence of contraband at that time or at sometime in the past, this observation takes too narrow a view of the process, ignoring as it does the virtual certainty that a full-fledged search will follow the alert, a search which, if the dog, or the handler's interpretation of the dog's behavior, was wrong, *will* reveal innocent contents and invade privacy. *See Caballes*, 125 S.Ct. at 842 (Souter, J., dissenting)(after positive alert, "[t]he police will then open the container and discover whatever lies within, be it marijuana or the owner's private papers"). This should rightly be a matter of concern, since, in our post-*Whren*, post- September 11, world, law enforcement officers can — at borders, at airports, at train stations, at bus terminals, on the highways — expose almost anyone they choose, or their vehicles or their possessions, to drug-sniffing dogs, either randomly or in a more targeted fashion, knowing that an "alert" will provide *all* the justification they need for a full-scale search of the individual and his or her effects. These facets of the *Caballes/Place* fallacy should be stressed at every opportunity, until courts evidence a more nuanced comprehension of what all the fuss is about than they, for the most part, presently do.

8. *See, e.g.*, United States v. Sanchez, 417 F.3d 971, 976 (8th Cir. 2005); United States v. Williams, 403 F.3d 1203, 1207 (10th Cir. 2005); United States v. Robinson, 390 F.3d 853, 874 (6th Cir. 2004); United States v. Williams, 365 F.3d 399, 406 (5th Cir. 2004); United States v. Cedano-Arellano, 332 F.3d 568, 573 (9th Cir. 2003), *cert. denied*, 540 U.S. 1137 (2004); United States v. Carter, 300 F.3d 415, 422 (4th Cir. 2002), *cert. denied*, 537 U.S. 1187 (2003); United States v. Ward, 144 F.3d 1024, 1031 (7th Cir. 1998); Karnes v. Skrutski, 62 F.3d 485, 498 (3d Cir. 1995); United States v. Banks, 3 F.3d 399, 402 (11th Cir. 1993), *cert. denied*, 510 U.S. 1129 (1994); United States v. Navedo-Colon, 996 F.2d 1337, 1339 (1st Cir. 1993); United States v. Glover, 957 F.2d 1004, 1013 (2d Cir. 1992). *But see* United States v. Trayer, 898 F.2d 805, 808 (D.C.Cir. 1990)(dog alert at door of train roomette did not by itself establish probable cause), *cert. denied*, 498 U.S. 839 (1990).

9. R.C. Bird, *An Examination of the Training and Reliability of the Narcotics Detection Dog*, 85 Ky. L.J. 405, 422 (1997)(hereinafter "Bird"), quoting United States v. Paulson, 2 M.J. 326, 330 n.5 (A.F.C.M.R. 1976), *remanded by* 7 M.J. 43 (C.M.A. 1979). *See also Outlaw*, 134 F.Supp.2d at 813 ("The reliability of the canine alert depends significantly on the ability and reliability of the human handler").

10. So much for granted do some courts take the reliability of a canine alert that in cases involving warrantless searches in the Fifth Circuit, where "probable cause is developed on site as a result of a dog sniff of a vehicle," no showing of the dog's training and reliability is required. United States v. Sanchez-Pena, 336 F.3d 431, 444 (5th Cir. 2003); United States v. Williams, 69 F.3d 27, 28 (5th Cir. 1995), *cert. denied*, 516 U.S. 1182 (1996). This stands on its head the rule that the burden is on the government to establish the constitutionality of a warrantless search, which the Fifth Circuit has recognized in other cases. *See, e.g.*, United States v. Rivas, 157 F.3d 364, 368 (5th Cir. 1998)(invalidating search where evidence was that dog "cast", i.e., stopped and paid close attention, but did not give the aggressive final alert for which it was trained). The burden should remain on the government to establish real probable cause for the search, which in this context should require, at a minimum, a showing that the dog and the handler were trained, that the dog actually alerted, and that that alert is a reliable indicator of the presence of drugs. *See, e.g.*, United States v. Swanger, 2005 WL 2002441 at *5-*6 (E.D.Ky. August 18, 2005)(government bears the burden of establishing the dog's training and the reliability of the dog's positive reaction; refusing to consider dog alert as basis for probable cause to search car where government did not establish dog's qualifications).

11. In United States v. Scarborough, 128 F.3d 1373, 1378 (10th Cir. 1997), for example, the dog had an overall reliability rate of 92%, but the dog's reliability rate when working with postal inspectors was only 79%. The court rejected the defendant's argument that it should look to the reliability rate when working with postal inspectors, as it had been in that case,

because no reason had been provided why the dog's abilities might be materially affected when working with postal inspectors. The lessened reliability when working with postal inspectors may well have been attributable to a lower level of handler training or experience with the reactions of that particular dog.

12. Bird, *supra* note 9, at 422.

13. Bird, *supra* note 9, at 412, 422-23.

14. *See, e.g.*, United States v. Sundby, 186 F.3d 873, 876 (8th Cir. 1999); United States v. Kennedy, 131 F.3d 1371, 1376-77 (10th Cir. 1997), cert. denied, 525 U.S. 863 (1998); see also United States v. Berry, 90 F.3d 148, 153 (6th Cir.)(affidavit sufficient where it referred to dog as a "drug sniffing" or "drug detecting" dog trained and qualified to conduct narcotics investigations), *cert. denied*, 519 U.S. 999 (1996); United States v. Daniel, 982 F.2d 146 (5th Cir. 1993)(affidavit sufficient where it said only that the dog was trained to detect the presence of controlled substances); United States v. Sentovich, 677 F.2d 834, 838 n.8 (11th Cir. 1982) (affidavit sufficient where it merely said that dog was trained in drug detection; rejecting argument that affidavit was insufficient because it did not also say that an experienced handler was with the dog); United States v. Meyer, 536 F.2d 963, 966 & n.4 (1st Cir. 1976)(affidavit sufficient where it said only that dog was trained and used by DEA agents in drug investigations); *but see* United States v. Cuevas, 1995 WL 382346 at *2 (N.D.Cal. June 19, 1995)(where affidavit contained no information regarding the dog's reliability, the alert information was "of little, if any, value in determining probable cause").

15. At one end of the spectrum is United States v. Outlaw, 319 F.3d 701, 704 & n.1 (5th Cir. 2003), in which the dog had four weeks training and then the dog and the handler had two more weeks training, and the alert occurred little more than a month after the dog was certified. The subsequent search produced a drug that the dog was not trained to detect, and it is clear from the underlying district court opinion — although one would never know it from the Fifth Circuit's decision — that the handler displayed at the evidentiary hearing a marked level of ignorance regarding the training process. *See Outlaw*, 134 F.Supp.2d at 811. *See also* United States v. Delaney, 52 F.3d 182, 188 (8th Cir.)(handler had 76 hours of training), *cert. denied*, 516 U.S. 878 (1995). At the other end are cases such as United States v. Navarro-Camacho, 186 F.3d 701, 704 (6th Cir. 1999), in which the dog and handler had trained together for 1,500 - 2,000 hours over the course of many years. *See also* United States v. Diaz, 25 F.3d 392, 394 (6th Cir. 1994)(dog and handler attended eight-week training course); United States v. Lingenfelter, 997 F.2d 632, 639 (9th Cir. 1993)(dog and handler had participated in 300 hours of training searches).

16. *See, e.g., Rosario-Peralta*, 199 F.3d at 562 (noting testimony that the handler must interpret the signal from the canine); *Outlaw*, 134 F.Supp.2d at 813 ("an alert is simply an interpretation of a change in the dog's behavior by a human handler"); *see also* United States v. Bartz, 2004 WL 1465780 at *5 (S.D. Ind. June 25, 2004)(handler's training includes learning how to recognize the changes in dog's behavior signaling the detection of narcotics). In United States v. Johnson, 323 F.3d 566 (7th Cir. 2003), the Court concluded that the district court had erred in disregarding the canine alert in assessing the existence of probable cause on the ground that the dog's handler had not testified at the suppression hearing, although another officer, who had observed the sniff from a distance, testified that the dog alerted. *Id.* at 567-68. The handler, the Court stated, "was not the only officer capable of interpreting [the dog's] behavior as alerting to the presence of drugs or drug-infested currency." *Id.* In fact, the dog's trained handler *is* often the only person capable of reliably interpreting whether or not there had, in fact, been an alert.

17. For a description of the canine training process, see Bird, *supra* note___, at 410-15.

18. Not only do dogs trained to alert aggressively respond differently from dogs trained to alert passively, but there also may be a range of different responses within each category:

> The dog trained to alert aggressively tries to contact the scent source (biting, scratching, penetrating, attempting to retrieve), while the dog that alerts passively does not try to contact the scent source but instead performs trained behavior (sitting, looking at the source, sniffing toward the source, looking at the handler).

Johnson, 323 F.3d at 567, quoting Sandy Bryson, **Police Dog Tactics** 257 (2d ed. 2000).

19. J.G. Aristotelidis, *Trained Canines at the U.S.-Mexico Border Region: A Review of Current Fifth Circuit Law and a Call for Change*, 5 **Scholar** 227, 230-31 (2003).

20. *Id.*

21. In *United States v. Walton*, 2004 WL 3460842 at *7 (M.D.Tenn. November 12, 2004), the defendant argued that the dog had not given its trained final response to the package before it was opened by law enforcement officers. The handler testified that the dog had done so, and the defendant offered videotapes of the dog alerting during training exercises, presumably obtained through a discovery request, to contradict the handler's testimony. The court compared the videotape of the stop to the training videotapes and concluded that the dog had in fact given at least one trained final response.

22. *See, e.g.*, *Navarro-Camacho*, 186 F.3d at 706-07; United States v. McLaughlin, 2005 WL 2087853 at *3 (D.Utah August 22, 2005); United States v. Hbaiu, 202 F.Supp.2d 1177, 1180 (D.Kan. 2002); United States v. Heir, 107 F.Supp.2d 1088, 1091 (D.Neb. 2000). Canine sniffs are not, however, typically filmed or recorded.
23. For example, in *Hbaiu*, the handler testified that the dog was a passive alerter, meaning that it would sit down when it detected drugs. The defendant contended that the handler had ordered the dog to sit, although the handler and other officers at the scene denied this. While the videotape captured the dog sitting down at the rear corner of the trailer, the audio portion had, the officers testified, "malfunctioned" at this point during the encounter. 202 F.Supp.2d at 1180.
24. For example, in United States v. Ludwig, 10 F.3d 1523, 1528 (10th Cir. 1993), the handler testified that he knew how his dog alerted and that the dog had done so on the challenged occasion. In *Diaz*, 25 F.3d at 394-95, the dog's handler testified that the dog alerted by barking, biting, and scratching, but occasionally would alert by coming to a standstill in order to scent more intently. This latter behavior is likely not a true alert. Similarly, in United States v. Trayer, 898 F.2d 805, 808 (D.C.Cir.), *cert. denied*, 498 U.S. 839 (1990), the handler testified that the dog had been trained as an aggressive alerter, but that, on this occasion, it froze and pointed to the defendant's train compartment "like a bird dog," which was the way it alerted on the majority of occasions. This, too, was a dog which was not doing what it had been trained to do. In *Bartz*, 2004 WL 1465780 at *5, the handler testified that, under controlled circumstances, the dog would alert by sitting and staring, but that it had "intermediate behaviors" on the "path to final response;" *i.e.*, the dog would stretch up on his hind legs and stare if the drug were concealed in a high place or lie down if the drugs were concealed in a low place, and that the handler's training included learning to recognize the changes in the dog's behavior that signaled the presence of drugs. The court concluded that the dog had alerted by stretching on his hind legs and "locking up" at the minivan's rear bumper. This was, however, not a trained final alert; it was an "intermediate behavior." The court also found that the dog had alerted two other times, during neither of which the dog gave his trained final response. *See also* United States v. Gregory, 302 F.3d 805, 811 (8th Cir. 2002) (defendant's passenger testified that the dog did not alert; on appeal, Court concluded that district court had not clearly erred in crediting handler's testimony that dog had alerted), cert. denied, 538 U.S. 992 (2003); United States v. Pore, 328 F.Supp.2d 591, 595 (D.Md. 2004)(court finds it unlikely that handler would misinterpret dog's behavior because handler and dog had worked together for a year).
25. 107 F.Supp.2d 1088 (D.Neb. 2000).
26. *Id.* at 1091. *See also* United States v. $67,220.00 in United States Currency, 957 F.2d 280, 285-86 (6th Cir. 1992)(Court found little probative evidence that dog had alerted to currency where two officers who were at the scene testified that they did not know that the dog had alerted until the handler told them so). *But see Outlaw*, 134 F.Supp.2d at 813 ("a canine alert is not always an objectively verifiable event").
In United States v. Jacobs, 986 F.2d 1231, 1233-35 (8th Cir. 1993), the search warrant affidavit stated that the dog had showed interest in the package, but omitted the fact that the dog had not actually alerted to the package. The Court concluded that this omission was in reckless disregard of the truth and refused to apply the *Leon* good faith exception.
27. *See* Bird, *supra* note 9, at 424. One court has noted that "[a] false alert can result from the handler's conscious or unconscious signals given by the handler that lead a dog to where the handler suspects contraband items to be located." *Outlaw*, 134 F.Supp.2d at 813. Such an alert can be "false" in two entirely different ways. First, it could be a false positive alert because no drugs were actually present. Second, it could be a false alert even if drugs were present, if what the dog was reacting to was the handler's cues rather than the odor of drugs. In the latter case, the alert was "false" because the dog was reacting to something other than that to which it was trained to react. A cued alert is not a reliable indicator of the presence of drugs because it is not a manifestation of the dog's trained response.
28. 898 F.2d 805 (D.C.Cir. 1990)
29. *Id.* at 809.
30. *Id.*
31. *Id. See also Diaz*, 25 F.3d at 396.
32. 2005 WL 2087853 (D.Utah August 22, 2005).
33. *Id.* at *3. The defense expert testified that the handler was very suggestive in soliciting a response; the government's expert testified that the handler had not induced the alert because he had used the same behavior all around the vehicle. The court simply assumed that if the handler's overly-suggestive conduct had influenced the dog, it would have falsely alerted at other areas of the car before alerting to the location where the drugs were found. *Id. See also* United States v. Limares, 269 F.3d 794, 797-98 (7th Cir. 2001)("We can't exclude the possibility that [the dog's] success is just a mirror of the agent's ability to find drug-laden packages to put under her nose; maybe she would not fare as well on a randomly selected sample, but that possibility was not pursued at the hearing").
In United States v. Burnett, 240 F.Supp.2d 1183, 1193 (D.Kan. 2002), the defendant

contended that the handler jerked on the dog's chain to cause it to alert after the dog initially showed no response; the handler denied that he had, and the court, as usual, credited the testimony of the handler. Similarly, in *Hbaiu*, the defendant contended that the handler directed the dog with a hand signal; the handler denied that he had, and no hand signal was visible on the videotape of the encounter. 202 F.Supp.2d at 1180.

34 107 F.Supp.2d at 1091. *See also* United States v. Stephenson, 274 F.Supp.2d 819, 824 n.1 (S.D.Tex. 2002)(noting that dogs may be entirely without bias, but their handlers may not be), *aff'd* 95 Fed. Appx. 604 (5th Cir.), *cert. denied*, 125 S.Ct. 139 (2004).

35. *Id.*

36 *Outlaw*, 134 F.Supp.2d at 813. *See, e.g.*, United States v. Brock, 417 F.3d 692, 696 (7th Cir. 2005("[T]he dog's error rate might affect whether a warrant issued in reliance on the dog sniff was supported by probable cause"); *Kennedy*, 131 F.3d at 1377 ("A dog alert might not give probable cause if the particular dog had a poor accuracy record"); *Ludwig*, 10 F.3d at 1528 (same); *McLaughlin*, 2005 WL 2087853 at *2 ("Although there may be circumstances that would bring the reliability of a specific narcotics dog or a specific dog search into question, those appear to be limited and rare). *See also Diaz*, 25 F.3d at 396 (a *very low* percentage of false positives is not necessarily fatal to a finding that a drug detection dog is properly trained and certified" (emphasis added)).

37. The Ninth Circuit, for example, has held that discovery of the dog's "qualifications," which appears to include, at a minimum, training and certification records, is mandatory. *Cedano-Arellano*, 332 F.3d at 571, 573. *See, e.g.*, United States v. Lambert, 351 F.Supp.2d 1154, 1162 (D.Kan. 2004)(granting defendant's motion for the dog's training and certification records for the year prior to the search); United States v. Schwandt, 2004 WL 1846126 at *4-*5 (D.Kan. May 7, 2004)(government agreed to provide training, testing, and performance records, but other requests denied by court); United States v. Wood, 915 F.Supp. 1126, 1133, 1136 (D.Kan. 1996)(in response to defendant's discovery motion, government agreed to furnish records of the dog's certification and recertification and records of its structured training over a reasonable period of time, but court declined to order any of the additional discovery sought by defendant), *rev'd on other grounds*, 106 F.3d 942 (10th Cir.1997). *See also Robinson*, 390 F.3d at 874 (court reviewed performance statistics and other documentation regarding dog); United States v. Owens, 167 F.3d 739, 749 (1st Cir.)(defense expert reviewed dog's training and performance records), *cert. denied*, 528 U.S. 894 (1999); *Kennedy*, 131 F.3d at 1374-75 (court discusses what the "discovery process" revealed about the dog and its trainer)

38. A dismal example of such misguided myopia is United States v. Gonzalez-Acosta, 989 F.2d 384, 388 (10th Cir. 1993), in which the defendant sought, by way of a Fed. R. Crim. P. 17(c) motion rather than a motion for discovery, discovery of the dog's training and veterinary records, false-positive/false-negative alert records, and "all other records establishing the dog's ability to smell." The Court, extraordinarily, essentially ruled that because the dog had been right in the instance before it, the records were not relevant:

> [W]e do not believe the documents were relevant because the dog was certified on the day in question and because the dog properly alerted to the presence of contraband. . . . Indeed, had the dog's records indicated it had false alerted in the past, defendant's ability to cross-examine would not have been enhanced because there is no doubt it correctly alerted in this instance.

Id. at 389. *See also McLaughlin*, 2005 WL 2087853 at *3 (past failure would not detract from the accuracy of the dog in this case). Whether drugs were found after the dog alerted in the case before the court is, of course, fundamentally irrelevant to the question whether the canine alert provided probable cause for the search. In the canine alert context, as in every other search context, probable cause is to be determined by the facts and circumstances as they existed *before* the search, not retroactively validated by the results of the search. *See, e.g., Pore*, 328 F.Supp.2d at 594 (must look at facts as they existed before drugs were found); *Outlaw*, 134 F.Supp.2d at 815 (whether there was probable cause at the time of the search may not be determined by what the search turns up).

The Sixth Circuit has repeatedly held that in order to prove the dog's reliability, the government need not provide the dog's training and performance records and that the testimony of the dog's handler will suffice for this purpose. *See, e.g.*, United States v. Boxley, 373 F.3d 759, 761 (6th Cir.), *cert. denied*, 125 S.Ct. 435 (2004); United States v. Hill, 195 F.3d 258, 273 (6th Cir. 1999), *cert. denied*, 528 U.S. 1176 (2000); *Diaz*, 25 F.3d at 394. It has acknowledged, however, that "[l]ack of additional evidence, such as documentation of the exact course of training, . . . would affect the dog's reliability." *Id.* The converse may also be true in any given case: examination of the underlying documentation may demonstrate that the dog is not in fact as reliable as its handler would have it. Moreover, as will be discussed in the text, *supra*, discovering that documentation does not exist may also be of significance.

39. A ridiculous example of this insistence that handler testimony is sufficient to establish the

dog's reliability is found in *Hill*, 195 F.3d at 273-74. In that case, the defendant argued, based in part on the handler's admitted failure to keep a record of the dog's false alerts, that the district court's finding that the dog was properly trained and reliable was clearly erroneous. The Court disagreed, noting that the dog's trainer had reviewed the dog's training and performance records and had opined, based on those records — which did not include records of the dog's false alerts — that the dog was reliable. *See also Diaz*, 25 F.3d at 395-96 (rejecting defendant's argument that reliability was not demonstrated because dog's training and performance records were not produced, and handler was unable to answer precisely questions regarding how many searches dog had done and how many times drugs were and were not present).

40. In United States v. Warren, 997 F.Supp. 1188 (E.D.Wis. 1998), the handler credited the dog with 100% accuracy. However, other evidence showed that when the dog was brought to a scene, the dog would alert to the suspected container, but generally only after some direction or coaching, and drugs might or might not be found in the container. If no drugs were found, the handler did not record it as a false positive alert but instead "note[d] that the dog must have smelled the residual odor of drugs which must have been present at some time in the past." *Id.* at 1192. *See also Gregory*, 302 F.3d at 811 n.1 (handler testified that on occasions when no drugs were found after dog alerted, there was evidence that drugs had been in the location a short time before the dog alerted); United States v. Martin, 2004 WL 2011456 at *3 & n.1 (D.Kan. August 19, 2004)(handler testified that dog had never falsely alerted, but this testimony was based only on training searches because, the handler said, only in such a controlled environment could he be sure that the odor of drugs was not present); United States v. Page, 154 F.Supp.2d 1320, 1325 (M.D.Tenn. 2001)(handler testified that, while there had been occasions when no drugs were found after the dog alerted, alerts to lingering odor of drugs are considered accurate alerts).

41. For example, in *Pore*, the defense presented the court with the dog's field reports, which counsel argued showed that drugs were found on only half the occasions when the dog alerted. The court found that, because there had been no testimony concerning the reports and no explanation of how they should be read or interpreted, they had "scant evidentiary value, particularly when a trained dog's sense of smell is powerful enough to detect a lingering odor of narcotics, even when the drugs are gone or there is no visible residue." 328 F.Supp.2d at 594 n.7. In United Sates v. Arreola-Delgado, 137 F.Supp.2d 1240, 1244-45 (D.Kan. 2001), the dog aggressively alerted to a suitcase which had been removed from the trunk of a vehicle, but no drugs were found in the suitcase. Upon further exploration, cocaine, some of which had leaked from its packaging, was found in the gas tank. The Court bought the handler's explanation that the alert could be explained by the wind's having blown over the gas tank, carrying the odor from underneath the car, and bouncing it off the nearby suitcase. *See also Page*, 154 F.Supp.2d at 1325 & n.1; *Outlaw*, 134 F.Supp.2d at 815; United States v. Fisher, 2002 WL 563581 at *8 (E.D.Pa. April 15, 2002).

42. In *Rosario-Peralta*, for example, no drugs were found on the vessel after the dog positively alerted to the deck; the defendants denied having been in the possession of cocaine and contested the testimony of law enforcement officers to having observed the vessel dumping bales of cocaine. The Court found the canine alert evidence admissible against a Rule 403 challenge because "the positive canine alert corroborated the officers' testimony by demonstrating that an allegedly unbiased canine indicated that a controlled substance contaminated defendants' vessel." 199 F.3d at 562. *See also Boxley*, 373 F.3d at 761 (dog alerted to defendant's pocket, but no drugs found; evidence admitted to connect defendant to drugs found nearby on theory that "an alert in the context of a canine narcotics sniff indicates that narcotics are present in the item being sniffed or have been present in such a way as to leave a detectable odor").

43. *See, e.g.*, United States v. Funds in Amount of $30,670.00, 403 F.3d 448, 460 (7th Cir. 2005)(noting that drugs or currency were found after 97.6% of the dog's alerts); *Limares*, 269 F.3d at 797 (62% of alerts followed by discovery of drugs, 31% followed by discovery of currency, 7% were unambiguous false positives); *Navarro-Camacho*, 186 F.3d at 704 (90-97% accuracy rate); *Scarborough*, 128 F.3d at 1378 (overall reliability rate of 92%); United States v. Hephner, 260 F.Supp.2d 763, 770 (N.D.Iowa 2003)(80% accuracy rate), *aff'd* 103 Fed. Appx. 41 (8th Cir. 2004). Unfortunately, even demonstrating that the dog has an abysmal accuracy rate frequently does not lead to a finding of no probable cause. *See, e.g., Limares*, 269 F.3d at 798 (62% accuracy rate suffices to demonstrate probable cause); *Kennedy*, 131 F.3d at 1378 (70-80% accuracy rate suffices to demonstrate probable cause). And such statistics may not in fact mean what the courts think they do. See R.E. Myers II, *In The Wake of Caballes, Should We Let Sleeping Dogs Lie?*: 20 **WTR Crim. Just** 4, 10-11 (2006)

44. In United States v. Cruz-Roman, 312 F.Supp.2d 1355, 1361 (W.D.Wash. 2004), for example, the court concluded that the dog/handler team was insufficiently reliable to be the basis for a finding of probable cause because "[r]ecords of training exercises and actual narcotics investigations indicate that the team often made mistakes and had no history showing a strong percentage of correct search results."

45. Bird, *supra* note 9, at 421-22. *See* United States v. Race, 529 F.2d 12, 14 (1st Cir. 1976)

(dog had four hours per week follow-up training).

46. *See Gregory*, 302 F.3d at 811 (noting that handler kept daily records of the dog's performance in ongoing training and in the field).

47. *Kennedy*, 131 F.3d at 1374-75. The handler in *Kennedy* ignored these directives, did not keep records of the dog's field work, and only field trained the dog sporadically. What records there were indicated an accuracy rate of 71.4%, but the handler's sloppy recordkeeping precluded assessment of the dog's actual accuracy rate based upon all its alerts. Unfortunately, the Court, referring to its general rule that search warrant affidavits are sufficient if they merely state that the dog was trained and certified and "declin[ing] to encumber the affidavit process by requiring affiants to include a complete history of a drug dog's reliability," *id.* at 1376-77, concluded that, even had all the information revealed at the suppression hearing regarding the dog's reliability and the handler's "sloppy conduct" been included in the affidavit, a reasonable magistrate judge would still have issued the warrant. In a case involving the same dog/handler team and a warrantless search, the court reached the right result, concluding that because of the inadequacy of the documentation of the dog's activities, the government had not demonstrated the dog's reliability and, hence, there was no probable cause for the search of the defendant's luggage. United States v. Florez, 871 F.Supp. 1411, 1422-23 (D.N.M. 1994).

48. They have not, with few exceptions, done so yet. In *Hill*, for example, the dog's handler testified that he did not know what ongoing training he was supposed to perform with his dog and that he did not keep records of false alerts. The Court, following circuit precedent, rejected the defendant's challenge to the government's failure to produce records to establish the dog's reliability. 195 F.3d at 273. In the earlier case, however, the Court, while concluding that the testimony of the dog's handler sufficed to establish the dog's reliability, did note that "training and performance documentation would be useful in evaluating a dog's reliability." *Diaz*, 25 F.3d at 396. *See Outlaw*, 134 F.Supp.2d at 811 (defense expert testified that the reliability of a particular dog can be determined only if records of both the dog's training and success and failure rate in the field are kept). *See also* note ___, *supra.*

49. *See* note 14, *supra.*

50. *See, e.g., Limares*, 269 F.3d at 798; *United States v. Klein*, 626 F.2d 22, 27 (7th Cir. 1980).

51. In United States v. Martinez, 78 F.3d 399 (8th Cir. 1996), the affidavit said nothing about the dog other than that it had alerted to the vehicle sought to be searched. It omitted to mention that the dog had been wrong in eleven of its prior twelve alerts. *See id.* at 401. The majority did not address the issue, concluding that the affidavit established probable cause without the alert. In dissent, Judge Arnold, addressing the issue because he disagreed that probable cause existed without the alert, stated flatly: "authorities do emphatically have a duty to inform the magistrate if a drug dog is unreliable." *Id.* at 403 (Arnold, J., dissenting). Omitting the dog's abysmal track record, he continued, constituted reckless disregard for the truth, and the good faith exception would not apply, because knowledge of the dog's unreliability made it objectively unreasonable for the officers to rely on the warrant. *Id. See also* United States v. Jackson, 2004 WL 1784756 at *3 n.1 (S.D.Ind. February 2, 2004)(questioning whether incomplete report about the dog's reliability was sufficient to support probable cause, but indicating that good faith exception would apply).

52. *See, e.g., Delaney*, 52 F.3d at188; United States v. Williams, 2000 WL 979997 at *7 (S.D.Ohio June 5, 2005); United States v. Cortez, 1995 WL 422029 at *3 (S.D.N.Y. July 18, 1995).

53. *See* Bird, *supra* note 9, at 425-26.

54. 438 U.S. 154 (1978).

55. *See, e.g., Limares*, 269 F.3d at 797-98; *Sundby*, 186 F.3d at 876; *Kennedy*, 131 F.3d at 1377; *Lingenfelter*, 997 F.2d at 640; *Jacobs*, 986 F.2d at 1234. n

National Association of Criminal Defense Lawyers (NACDL)
1660 L St., NW, 12th Floor, Washington, DC 20036
(202) 872-8600 • Fax (202) 872-8690 • assist@nacdl.org

DECLARATION OF SERVICE BY MAIL

Re: People v. Bailey Lamar Jackson No. S139103

I, RICHARD I. TARGOW, certify:

I am, and at all time mentioned herein was, an active member of the State Bar of California and not a party to the above-entitled cause. My business address is Post Office Box 1143, Sebastopol, California 95473.

I served a true copy of the attached APPENDIX TO APPELLANT'S OPENING BRIEF on each of the following, by placing same in an envelope or envelopes addressed, respectively, as follows:

Office of the Attorney General
P.O. Box 85266-5299
San Diego, CA 92186-5266

Steven Parnes, Staff Attorney
c/o CAP Docketing Clerk
California Appellate Project
101 2nd Street, Suite 600
San Francisco, CA 94105

Hon. Patrick F. Magers,
c/o Clerk of the Superior Court
P.O. Box 431,
Riverside, CA 92501

Bailey L. Jackson, Jr. (Appellant)

Each said envelope was then, on June ___, 2012, sealed and deposited in the United States Mail at Sebastopol, California, with postage fully prepaid. I declare under penalty of perjury that the foregoing is true and correct.

DATED: June ___, 2012

RICHARD I. TARGOW
Attorney at Law