Applied Animal Behaviour Science 284 (2025) 106527

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim

The effect of training paradigm on dogs' (*Canis familiaris*) acquisition and generalization of smokeless powders

S.A. Kane [a,*], G.N. Cupp [a], M. Rangel [b], A. Medrano [b], J. Davis-Miller [c], C. Collins-Pisano [c], B. Rogers [c], D. Copeland [c], L. Lazarowski [c], P. Waggoner [c], E.O. Aviles-Rosa [a], P.A. Prada-Tiedemann [b], N.J. Hall [a]

[a] *Canine Olfactory Research and Education lab, Davis College, Texas Tech University, Lubbock, TX, United States*
[b] *Forensic Analytical Chemistry and Odor Profiling Laboratory, Department of Environmental Toxicology at Texas Tech University, Lubbock, TX, United States*
[c] *Canine Performance Sciences, Auburn University College of Veterinary Medicine, Auburn, AL, United States*

ARTICLE INFO

Keywords:
Dog
Explosives Detection
Generalization
Olfaction

ABSTRACT

Explosives detection canines need to be trained to detect a range of energetic variations at an efficient scale to meet the needs of public safety. The goal of this study was to determine which of three training paradigms most efficiently trained canines to detect eight smokeless powder (SP) varieties and to generalize to novel SP exemplars. Three training paradigms were compared: Sequential (dogs were trained to one odor at a time), Mixture (dogs were trained to a mixture of four SPs), and Inter-mixed (dogs were trained to four unmixed SPs within a session). SPME GC-MS was used to evaluate the volatile organic compound (VOC) profile of 18 SPs. We then identified four SPs that showed maximal VOC profile variation as the initial training targets and four related variants. The training paradigm had no effect on acquisition time. One anomalous SP was observed, where 12 of 18 dogs failed to reach criterion. Inter-mixed training led to higher rates of generalization across 10 untrained SP varieties compared to Sequential and Mixture training. Mixture and Sequential trained dogs did not differ in their generalization rates. Although Inter-mixed training led to higher rates of generalization, it did not produce proficiency in detection of all novel variants, with many variants (double and single base) producing < 75 % response rates. Explicit training with some variants is still required with Inter-mixed training. A Partial Least Squares Regression (PLS), where the SPME GC/MS VOC peak areas for each detected VOC were used to predict canine nose hold time, explained 72 % of the variance in VOC data and 17 % of the variance in cumulative nose hold time. The VOCs with the highest variable importance was one VOC unique to Vihta Vuori® single bases, and two VOCs associated with Hodgdon® single bases, to which dogs showed poor generalization. Good to adequate generalization was observed across Hodgdon® double bases and Accurate® single and double bases. Overall, these results suggest that Inter-mixed training has important benefits in generalization; however, explicit training is needed to reach proficiency.

1. Introduction

Smokeless powders (SP) are one of the most used propellants for improvised explosive devices (IEDs) within the United States (Bureau of Alcohol, Tobacco, Fireams and Explosives, 2022), and explosives detection canines (EDCs) remain the best real-time threat detection tool for these devices (Furton and Myers, 2001). One challenge for detection dogs is the need to respond to a wide range of possible variations of an energetic material. Generalization is a key skill for detection canines (Aviles-Rosa et al., 2021; Hall and Wynne, 2018; Lazarowski and Dorman, 2014). Failures of generalization pose a public safety risk, and thus it is crucial that we develop training methodologies that increase the likelihood of generalization without needing to train to all possible targets in the field (Keep et al., 2021).

SPs can be classified as single base (nitrocellulose base charge), double base (nitrocellulose and nitroglycerin base charges), or triple base (nitrocellulose, nitroglycerin and nitroguanidine base charges) (Lennert and Bridge, 2018). Chemical analysis of different SPs showed that the headspace of all SPs tested contain 2-ethyl-1-hexanol, diphenylamine (DPA) and dibutyl phthalate in their headspace (Rangel et al.,

* Corresponding author.
E-mail address: sarahkan@ttu.edu (S.A. Kane).

https://doi.org/10.1016/j.applanim.2025.106527
Received 28 October 2024; Received in revised form 10 January 2025; Accepted 16 January 2025
Available online 30 January 2025

in prep). Even though SP variants share similarities in headspace compounds, detection dogs do not always spontaneously generalize to different SP types (Aviles Rosa et al., 2022). Previous literature indicates that certain variants require additional training (Oxley and Waggoner, 2009). One potential reason for these generalization failures is that SPs can contain different additives, which change their overall odor profile (Lennert and Bridge, 2018).

Dogs can be trained to detect multiple target odors using a variety of methods. One of the most common methods is sequentially, where dogs are trained to detect one odor at a time, advancing to the next odor only after meeting a pre-determined criterion. Another method, mixture, mixes several different odors and trains the dogs to the mixed odor. A third method, inter-mixed, trains dogs to multiple odor variations in discrete trials simultaneously within the same sessions.

There is mixed support for some of these methods. For example, Gazit et al., (2021) demonstrated that dogs trained to detect a three or a five-component mixture generalized to the individual explosive components (Gazit et al., 2021). Due to security concerns, however, the authors could not disclose the explosives used in these mixtures, limiting applicability because odor identity can have an important impact on mixture perception.

Keep et al., (2021) directly compared sequential, mixture and inter-mixed training methods in rodents and found that mixture training reduced overall training duration compared to the other two training methods, but that inter-mixed training increased generalization to novel targets compared to the other two methods.

Comparisons of these three training methods have not been empirically evaluated in dogs using explosives as target odors. The goal of this experiment is to compare training duration and generalization for these three training paradigms (Sequential, Mixture and Inter-mixed) in detecting the highly variable explosive class of SP. This study was done in conjunction with another study, which explored these same paradigms in training multiple explosives classes (plastics, AN varieties, TNT varieties, PETN-based; Kane et al., in review).

With SPs, we expected that Mixture and Inter-mixed training methods would result in the shortest training duration. SPs are from one explosive class suggesting similar salience of each component of the mixture reducing the potential of overshadowing (Thomas-Danguin et al., 2014). We hypothesized that an odor mixture containing SPs that maximize the variation in VOC profiles, may simultaneously train dogs to all relevant SP odor features. We also expected Inter-mixed trained dogs to exhibit higher generalization to novel targets based on prior research in rodents (Keep et al., 2021).

2. Methods

2.1. Participants

Eighteen dogs participated; twelve dogs were from the Texas Tech Canine Olfaction Research and Education Lab (TTU) and six dogs were from the Auburn University Canine Performance Science (AU) program. The AU dogs were purpose-bred working canines, and the TTU dogs were selected from a local animal shelter for a training for adoption program based on their food motivation and trainability (see Table S5 for detailed demographic information). All procedures were approved under TTU IACUC protocol number: 2022–1180 and AU IACUC protocol number: 2022–5029.

Dogs were trained in cohorts of six (TTU) and three (AU). One TTU cohort was previously naïve to explosive odor detection and the other had previously participated in a related experiment for training different classes (Kane et al., in review). All AU dogs were experienced in odor detection but were naïve to the explosives used in this experiment. Like the TTU cohorts, the first AU cohort had completed the related between class explosive study prior to the current experiment and the second AU cohort had not. Cohorts were started on different experiments to limit order effects. Dogs were kept in the same experimental training paradigms across the two experiments.

2.2. Experimental design

Dogs were trained in the first phase (acquisition) using the three training paradigms in an automated three-olfactometer scent line-up. Once these dogs met criterion on eight SP varieties, post-acquisition generalization was evaluated to an additional ten novel SP varieties.

Eight SPs were used as initial training targets. The first four of these targets (primary targets) were selected to represent the entire odor VOC space based on a chemical analysis PCA (details described below). The four primary target SP varieties were double based Hodgdon® US869, Hodgdon® CFE 223 and Accurate® 1680, and single base Vihta Vuori® N165. Another four targets were selected as generalization probe odors (SP variants) that overlapped a similar VOC PCA space to one of each of the four primary targets. These SP varieties were double base Hodgdon® H335, Accurate® 2200, and Vihta Vuori® 530, and single base Accurate® 2015. This yielded the eight total odors which were used to evaluate acquisition.

During the acquisition phase, generalization to the initial eight odors was assessed each time a dog met criterion on odor(s) they were training to detect. We conducted these generalization tests to allow dogs to "skip" explicit training on odors to which they spontaneously generalized and thus decrease overall session number to acquire all eight targets. Dogs were subsequently explicitly trained to any odors they did not generalize to on the generalization test. This train-generalization test-train cycle was repeated until dogs met generalization criterion for all eight SP varieties (see Fig. 1).

2.3. SP variety selection

Eighteen commercially available smokeless powders were purchased. SPME followed by GC/MS was conducted for each powder and reported in detail in Rangel et al. (in prep). The peak area for each VOC identified was compiled into a matrix for six replicate analyses for each of the eighteen smokeless powders. Principal component analysis was used to describe the VOC variation in two dimensions. PC1 explained 12% of the variation and PC2 explained 10% of the variation (see Fig. 2). To identify training samples that represented this entire "VOC space", we randomly selected one training SP and one probe SP from each of the quadrants from the PCA. The powders selected for training and probe testing in each quadrant are labeled in Fig. 2. The purpose was to select a series of four SP varieties that maximally represented the VOC variability space. The four generalization probes were selected to represent a similar PCA space as one of the primary training varieties. The SP not tested/trained were used for the post-acquisition generalization test.

The order of training the four primary targets and their associated variant was randomized by cohorts. Three of the four primary targets were trained first for one of the cohorts (the two cohorts from AU started with the same odor). Vihta Vuori® 165 was never in the first position of the primary targets. The order for variants (odors five to eight) followed the order for the related primary target. For description of odor order see Table 1.

2.4. Training paradigms

Three training paradigms were compared: Sequential, Mixture, and Inter-mixed (see Fig. 1). Dogs were assigned to one of the three training paradigms in a stratified random assignment at the start of the study or at the start of the associated study (between class explosive generalization). In each cohort an equal number of dogs were assigned to each training paradigm, so that in a cohort of six dogs there were two in the Sequential, two in the Inter-mixed and two in the Mixture group. AU and TTU had equal number of dogs in each training paradigm.

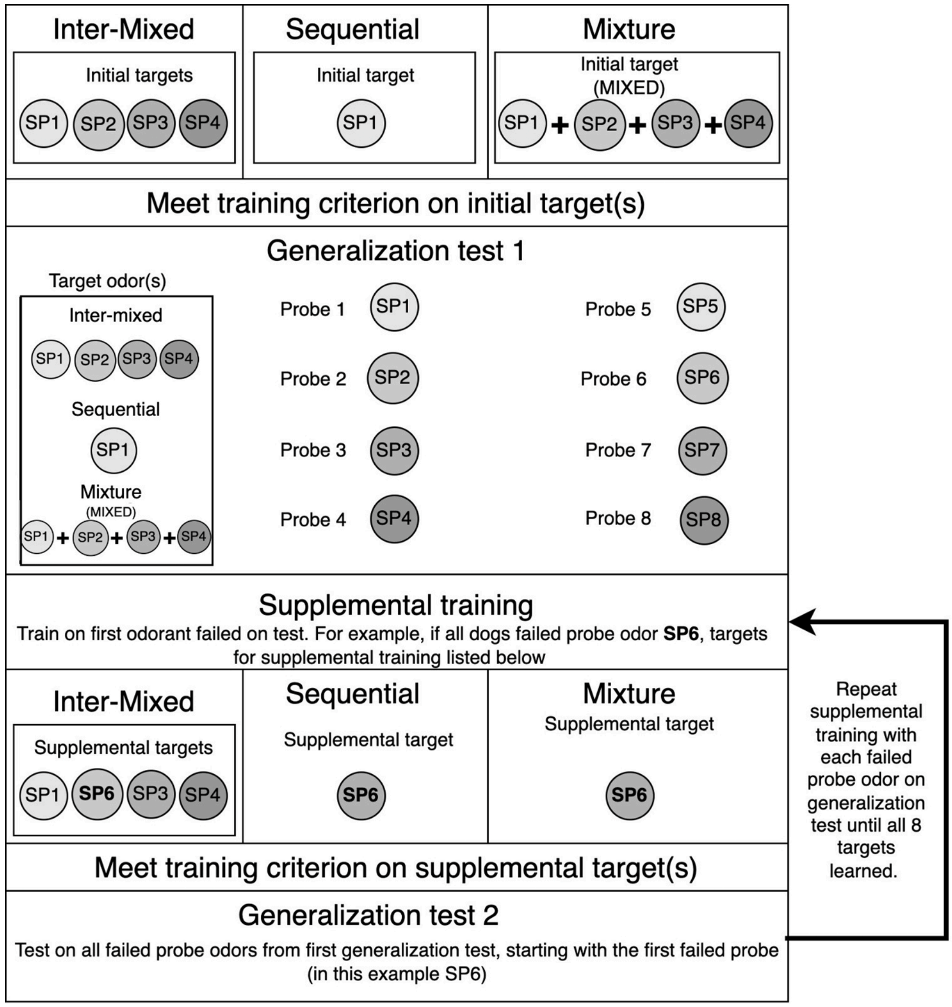

Fig. 1. Training and testing cycle diagram. This diagram shows how each of the three training paradigms learned the SP varieties during the acquisition phase.

2.4.1. *Sequential training*

Dogs were trained to detect one smokeless powder at a time. Once dogs met acquisition criterion to one powder, generalization was tested to all eight varieties. The dog would then learn the next SP to which they failed to generalize. SPs were trained in the order assigned to their cohort (for order details see Table 1). Dogs could continue training and generalization testing until meeting criterion on all eight SP varieties.

2.4.2. *Mixture training*

Dogs were presented with an equal parts mixture of the headspace of the four primary target materials until reaching the training criterion, at which point generalization testing was conducted for all eight SP varieties. If a dog failed to generalize to a component of the target mixture or probe odor, the dog was explicitly trained to that target odor presented alone (like Sequential training). After this first generalization test, dogs were trained to individual missed SP varieties until meeting the training criterion for all eight odors.

2.4.3. *Inter-mixed training*

Each of the four primary targets were presented within a session but on separate trials and in a randomized order across trials. Training continued until reaching the training criterion on each of the four targets, at which point generalization testing was conducted for all eight SP varieties. When a dog did not meet criterion on a probe/variant, the primary odor of that associated probe was substituted for the missed odor (see Fig. 1). Training was then continued until reaching criterion on the failed odor. Once meeting criterion with that odor, generalization to all remaining odors was reassessed.

Training followed this cycle until the dog met the generalization criterion for all eight odors. Dogs in the Inter-mixed training paradigm all started with the four targets weighted such that each would appear in blocks of three consecutive trials. The order in which these blocks were presented was randomized (i.e., the order of targets). After meeting 90 % accuracy, odors were weighted to only appear in randomized blocks of one (i.e. 1:1:1:1 ratio).

2.5. *Equipment*

Three automated olfactometers were used to present target odors and other distractor odors to the dogs. These three olfactometers were independent replicates of each other, with each olfactometer containing the same 12 odors. The olfactometers were controlled by and communicated with a computer via Bluetooth. A computer program randomized the odors presented in each box, with at most one olfactometer presenting a target odor each trial.

During each trial, air flowed from an air pump through a charcoal filter, which removed odor contaminants present in the air. This air was then split into two lines (odor line, dilution line); the odor line flowed through a rotameter and into a valve manifold. Air would then flow through the open valve(s). Valves were opened and closed based on instructions from the computer program. Air flowed through the open valve into another rotameter, which measured the specific flow rate. This secondary rotameter allowed for greater manipulation of odor concentration in a mixture. Air flowed from this secondary rotameter

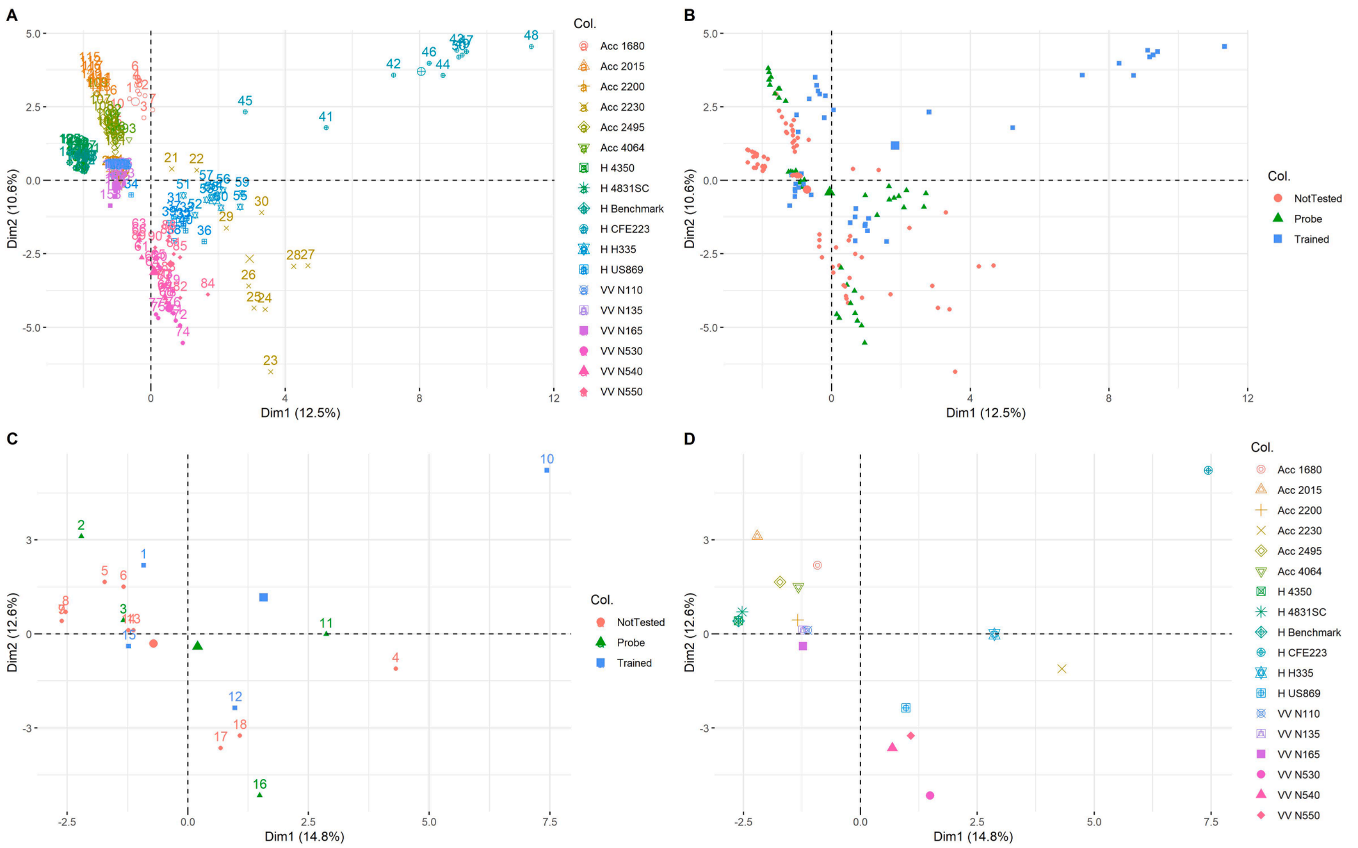

Fig. 2. PCA Analysis of Smokeless Powders. A: Shows the data including each SPME GC/MS replicate for each of the powders. B. Shows the same PCA but colored based on the four odors for explicit training and probe odors. Large geometric shapes show the mean for that category. C: PCA following taking the mean of all replicates for an odor. Shape indicates whether the powder was trained, probe or a variety not tested. D: Shows the same as C, but identifying the SP variety.

Table 1
Details of the smokeless powders used as the initial 8 target odors and variants in this experiment. Each cohort of six canines (combining the two cohorts of three from AU) were presented the initial four odors and their four variants in different orders.

SP Identifier	Brand	Base	Type	TTU Order Cohort 1 (n = 6)	TTU Order Cohort 2 (n = 6)	AU Order Cohort 1 (n = 6)
165	VihtaVuori®	Single	Initial target	3	2	4
US869	Hodgdon®	Double	Initial target	1	4	2
CFE223	Hodgdon®	Double	Initial target	4	3	1
1680	Accurate®	Double	Initial target	2	1	3
2015	Accurate®	Single	Variant of Accurate 1680	6	5	7
H335	Hodgdon®	Double	Variant of Hodgdon CFE223	8	7	5
2200	Accurate®	Double	Variant of VihtaVuori 165	7	6	8
530	VihtaVuori®	Double	Variant of Hodgon US869	5	8	6

into a glass vial containing an odor, the same volume of air pushed into the vial was displaced and flowed through a Teflon tube into the mixing manifold. Air from the dilution line would then mix with the odor line in the mixing manifold before it flowed to the odor port where it would be sampled by the canine (see Fig. 3).

All distractor odors were presented at a 1:1 dilution with clean air (i.e., 50 % dilution) at a total flow rate of 2 standard liters per minute (SLPM). For Sequential and Inter-mixed training all odors were diluted with a continuously running dilution line. For Mixture training the target was not diluted. The target was presented at a flow rate of 2 L/min, with each component of the mixture contributing 0.5 SLPM to the target mixture.

The olfactometers, in addition to presenting odors, calculated the duration each dog spent sniffing each odor via infrared (IR) beam pairs at the odor port. Based on pre-set alert times the computer program would sound a correct tone if the dog had correctly identified the target box and held an alert (nose hold in odor port) for the required alert duration. If the dog did not correctly hold an alert (no nose hold for required alert duration) or alerted to an incorrect olfactometer a different trial tone would play, indicating to the dog and the handler that the dog was incorrect and to not give reinforcement. In addition to these two sounds, there was a trial start tone and a trial end tone.

Each olfactometer contained four targets, at least one empty vial (blank control), and four distractor odors. There was also a spot for a probe odor for generalization testing. The distractor odors were replaced approximately once a week to ensure dogs were learning the target odors specifically, not learning to alert to any novel odor. Each generalization test also presented two novel distractors.

2.6. Initial olfactometer training

All dogs received similar initial training to the olfactometer. TTU

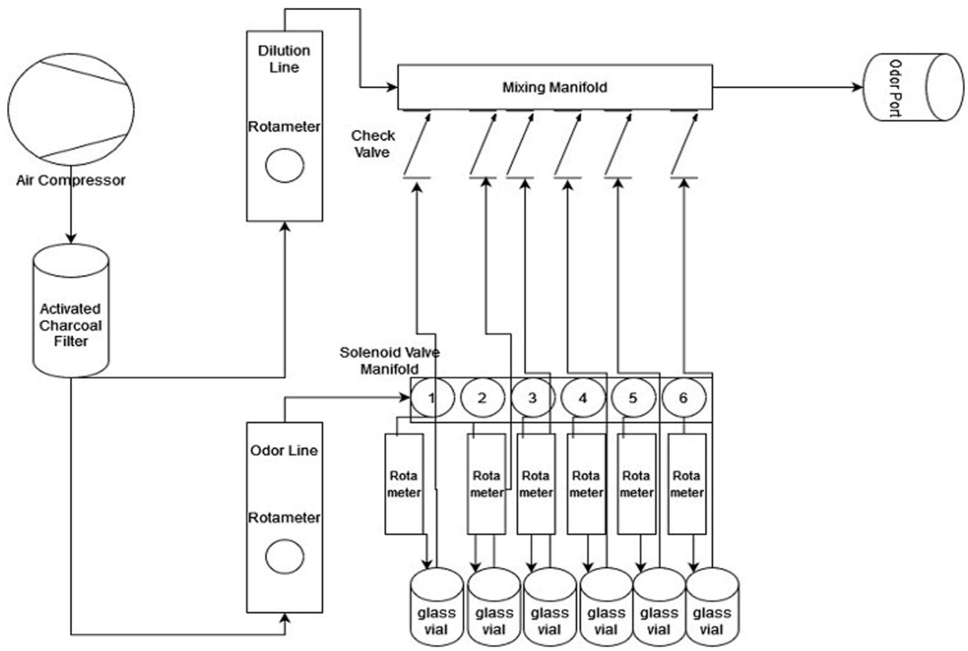

Fig. 3. Illustration of airflow through olfactometer. Note the olfactometer illustrated here only has six valves and odors for illustrative purposes, although the ones used in this experiment were expanded to 12 valves. Airflow indicated by arrow.

dogs were initially trained to detect food (hot dogs) in the olfactometers to teach initial search behaviors. Once dogs were holding a 1 second nose hold in the odor port for a food target they began the same initial training as the AU dogs. Dogs were first trained to detect a 1 % solution of iso-amyl acetate and then a 1 % solution of 1-bromo-octane in the olfactometers prior to acquisition training for SPs. In this initial training, dogs learned to perform a four second nose hold response, search the three olfactometers independently from the handler, perform "all clear" behaviors on non-target trials (20 % of all trials), and work at a 70% reinforcement rate. On non-target trials, the computer scored a correct response if the dog inserted their nose into each olfactometer (breaking the IR beam pair) and did not re-check olfactometers (stopped searching) for ten seconds. Also, during initial training, dogs were trained to ignore five to ten non-target distractor odors such as coffee. Distractor odors were used throughout the remainder of the experiment (see supplemental material Table S6). This training was done to ensure that dogs were familiar and comfortable with the experimental design and to ensure that dogs could perform under reduced reinforcement (necessary for generalization testing).

Table 2
Training procedures at each step during acquisition training.

Step	Alert Time (s)	Wait for Correct	Blinded	Non-target trials (% of trials)	Intermittent Reinforcement (% of trials reinforcers were available)	Accuracy Criterion
1	.1	Yes	No	0 %	100 %	None (2 sessions)
2	1	Yes	Yes	0 %	100 %	90 %
3	0.1–4	Yes	Yes	20 %	100 %	90 %
4	4	No	Yes	20 %	100 %	90 %
5	4	No	Yes	20 %	80 %	85 % (on 2 consecutive sessions)
6	4	No	Yes	20 %	70 %	85 % (on 2 consecutive sessions)

2.7. Training procedures

Dogs in each cohort were assigned into the Mixture, Sequential and Inter-mixed groups in a stratified random assignment, ensuring that at least two dogs per university cohort were assigned to each respective training paradigm (see Table S5 in supplemental materials). Experimental training for acquisition to SPs was developed to be as standardized as possible across groups to measure acquisition (see Table 2). All training sessions were 20 trials.

We changed distractor odors approximately every week for the duration of the experiment. See Table S6 (supplemental information) for a list of the distractors used.

The goal of the first stage of acquisition training (step 1) was to quickly reinforce the dog for sniffing the SP target, without requiring a change in behavior. During this stage, the alert duration was 0.1 s and there were no non-target trials. To encourage dogs to search all boxes without consequences we set all trials to "wait for correct," so the dog was reinforced for sniffing the target olfactometer but not penalized for sniffing other olfactometers. The handler was aware of the location of the target. The program would play a trial start tone, trial end tone and a correct tone, once the dog selected the target olfactometer. Dogs completed two sessions of 20 trials at this stage.

In the next training stage (step 2), we increased the alert duration to 1 second. All other settings were identical to step 1. Dogs were required to meet a 90 % accuracy on their first response on one session of 20 trials to move onto the next step.

The goal of step 3 was to introduce non-target trials and increase canine alert duration. We introduced four rewarded non-target trials (20 % of trials) where the dog was required to check all three olfactometers and leave the search area (i.e. make an "all clear"). Alert duration on this step randomly varied between 0.1 to 4 seconds. Dogs were only permitted two incorrect responses with an alert duration of 0.5 seconds or greater to move to the next stage. The program still "waited for the correct" response during this stage. The handler was blind to the location of the target from step 3 through 6.

In step 4 the alert duration was increased to 4 seconds, and the olfactometers no longer "waited for the correct response." An additional incorrect tone was added during this step. When dogs performed an "all clear" behavior when the target was present, or if they indicated on an olfactometer containing a distractor odor, the incorrect tone and trial end tone would play, and no reinforcer was given. Dogs were required to achieve 90 % accuracy at this stage to move to the next step.

Steps 5 and 6 trained dogs under reduced rates of reinforcement. On step 5, 80 % of target and non-target trials were reinforced. On step 6, 70 % of trials were non-reinforced. These non-reinforced trials were ambiguous, regardless of the dogs' response (correct/incorrect) the trial end tone played, and the trial terminated. Dogs were required to achieve 85 % accuracy across two consecutive sessions on both steps 5 and 6 to move to the next stage.

If a dog failed to meet criterion on a training step after five sessions at a single step and responded to the target odor on less than 50 % of the trials presented, that odor was removed from future training. This criterion was put in place after dogs were observed to be avoidant of one target (Vihta Vuori® 165). Dogs would begin a nose hold on this target but would remove their nose before the 4 second criterion. Vihta Vuori® 165 was the only odor for which this occurred. Dogs completed at least one generalization test with Vihta Vuori® 165 even if they failed to meet criterion. Twelve of eighteen dogs had Vihta Vuori® 165 dropped from their training odors.

2.8. *Generalization testing during acquisition*

Generalization testing determined if a dog adequately responded to an odor (explicitly trained or not), negating the need for further explicit training to that target. The canines were tested on all eight smokeless powder targets during their initial generalization test (the four primary targets and four variants). Two novel distractor odors were added for each generalization test, to measure alerting to novel distractor odors.

Generalization testing consisted of 20 trials. Sixteen trials were identical to the training in step 6 of the respective training paradigm and four trials were non-reinforced probe trials, where the novel target odor was presented in one olfactometer and distractors presented in the other two olfactometers. The location of these probe trials was randomized within the session. Of the 20 trials, two were non-reinforced trials to the target, one of which was a non-reinforced non-odor trials, and four non-reinforced probe trials to the odor being tested. Fourteen trials were reinforced, yielding a similar overall reinforcement rate as step 6. The probe trials were used to measure spontaneous generalization without explicit reinforcement. During probe trials, only the trial end tone would play, regardless of the dog's response.

An odor was considered successfully trained if the dog correctly alerted to that odor on three or more of the four probe trials in a session (binomial test, assuming chance $=0.33$, $p = 0.10$). That odor was then removed from further training and testing for that dog. If a dog did *not* alert on at least three of the four probes, the dog would continue training and testing on that odor until the three out of four correct probe trial criterion was met. A criterion of three of four was selected to yield a low probability of meeting by chance (10 % assuming a one out of three chance of alerting), but to not require perfect spontaneous generalization (4/4 alerts on probe trials).

If a dog did not meet the three of four criteria, dogs received "supplemental training" (described below) for these targets. If a dog previously received explicit training and passed step 6 for that odor, the dog started supplemental training at step 6, to re-meet step 6 criterion for that odor. That odor was then tested again. If an odor was not previously trained explicitly (e.g., a target variant or untrained primary target), supplemental training started at step 1.

If a dog required supplemental training for more than one odor following a generalization test, supplemental training occurred for the next odor in the odor assigned to that cohort (See Fig. 1 for diagram). Following meeting criterion for the new odor, generalization testing was repeated for all odors that did not previously meet criterion.

2.9. *Supplemental training*

Supplemental training was used to train odors that dogs did not meet criterion during the generalization test. This followed a similar progression through the six steps with minor differences between groups.

For Mixture training, each odor failed during the generalization test was then trained individually and sequentially (rather than as a new mixture), identical to the sequential group. Thus, these two groups only differed in the initial training to a single odor or the 4-component mixed odor. If a Mixture trained dog failed to meet the three out of four-probe response criterion on the generalization test to a component of the mixture target, the dog was explicitly trained with that odor in supplemental training (starting at step 1).

For Inter-mixed training, the first failed variant odor replaced one of the four initial odors. The new odor was weighted to appear more frequently (3x as frequent as the others) to facilitate training to the new odor.

2.10. *Post-acquisition generalization*

After the dogs had shown proficiency detecting the eight SP varieties, canine generalization was then tested with ten additional SP variants that had not been trained or evaluated during acquisition. The goal of this test was to determine if one training paradigm led to greater generalization to a wider range of novel SP variants.

One limitation of generalization testing during acquisition was that dogs in Inter-mixed training received multiple types of target odor trials during the session (due to the nature of their assigned training paradigm) whereas dogs in Sequential and Mixture training only had one target within the session. This difference may have led to greater generalization for Inter-mixed trained canines due to the difference in target variation *within* the generalization test session but may not have a broader impact on overall generalization. To overcome this limitation, post-acquisition generalization testing was conducted where all groups had the same single target (a trained target within the same PCA quadrant as the probe) odor presented within the generalization test session (see Table 3) and therefore any generalization differences would reflect their prior odor training history and not differences in the assessment procedure.

2.10.1. *Post-acquisition generalization: Training*

Following completion of the acquisition phase, and immediately before testing, dogs received refresher training using step 4 settings for their experimental paradigm in acquisition. Dogs were required to achieve a score of 90 % or greater across two 20 trial sessions on their initial target odor (i.e., the first trained target odor for the sequential group, the mixture target for the mixture group, and the inter-mixed targets for the inter-mixed group).

2.10.2. *Post-acquisition generalization: Testing*

The generalization test was identical for all dogs across the training paradigms. For each session, a singular trained SP target served as the target odor and a singular SP variety was evaluated in four probe trials. The most closely related (similar VOC profile) smokeless powder from acquisition training served as the target for each variant. SP variants were tested in a re-randomized order for each cohort.

Each day of generalization tests started with an initial refresher session to ensure performance to the target was satisfactory. If a dog scored greater than 80 % across ten trials, they were moved into probe testing for the day. Any score lower than 80 % resulted in a re-test of another ten trials. No dog needed to be re-tested more than twice.

The generalization test otherwise had identical settings to the generalization tests used during acquisition. Previously un-trained/un-tested distractor odors were included for each day of testing. One to three generalization tests occurred per day.

Table 3
Smokeless powder variants and odor order for post-acquisition generalization testing. This table details the brand, base, and type of smokeless powders used. This table also features the order of the smokeless powders presented to each cohort.

SP	Brand	Base	Type	TTU Order Cohort 1 (n = 6)	TTU Order Cohort 2 (n = 6)	AU Cohort 1 (n = 6)
2230	Accurate	Double	Variant of Hodgdon US869	10	1	5
550	VihtaVuori	Double	Variant of Hodgdon US869	9	2	4
540	VihtaVuori	Double	Variant of Hodgdon US869	8	3	3
N135	VihtaVuori	Single	Variants of Accurate 2200	7	4	2
N110	VihtaVuori	Single	Variants of Accurate 2200	6	5	1
Benchmark	Hodgdon	Single	Variant of Accurate 2015	5	6	10
H4350	Hodgdon	Single	Variant of Accurate 2015	4	7	9
2495	Accurate	Single	Variant of Accurate 2015	3	8	8
4831SC	Hodgdon	Single	Variant of Accurate 2015	2	9	7
4064	Accurate	Single	Variant of Accurate 2015	1	10	6

2.11. *Statistical analysis*

All statistical analyses were conducted in R version 4.3.2. Data were initially cleaned using dplyr and tidyverse packages (Wickham et al., 2019). Plots were created using ggplot2 (Wickham, 2016). ANOVA analyses were done using the car (Fox and Weisberg, 2018), lme4 (Bates et al., 2015), and lmerTest (Kuznetsova et al., 2017) packages. Post-hoc analyses were conducted using the emmeans package (Lenth, 2023).

We compared training duration (number of sessions required to meet criteria) by paradigm with a linear model with a random intercept for institution. We used mixed-effect models for analysis, in which we fit a random effect of institution to balance differences based on location of cohort. Dog was nested within institution. We used a logistic binomial mixed effect model in which training paradigm predicted probability of an alert to determine the effect of training paradigm on probability of an alert during post-acquisition testing. We analyzed the relationship between sniff time and odor identity on post-acquisition testing using a linear mixed effects model. Prior to running this analysis, we corrected all nose hold times greater than 4 seconds to equal 4 seconds. Some dogs with narrower-muzzles, or who moved more during their alert, would have greater cumulative sniff times due to repeatedly breaking the IR beam. A continuous 4 second nose hold was required so each time the muzzle shifted the timer would reset; artificially inflating nose hold time. Additionally, we removed all nose hold times of 0 seconds, as this value only represented that the dog did not check the odor port and therefore was not a measure of dog response.

To evaluate whether specific VOCs were related to dogs' generalization we used partial least squares (PLS) regression where the peak area for each of the 48 most frequent VOCs across all SPs was used to predict dog nose hold time for each SP. We combined all training data from Steps 4,5,6 and generalization data during and post-acquisition. This yielded detection performance data from 9582 trials spanning 18 SP varieties and 18 dogs. PLS regression was used to evaluate whether the peak area values from the 48 most frequently observed VOCs would yield a successful prediction of nose hold time. Cross validation was used to identify the optimal number of components. To enhance data visualization, we calculated a relative peak area index, where the peak area for each VOC was divided by the maximum peak area observed across the 18 SPs for that VOC.

3. Results

3.1. *Effect of training paradigm on acquisition*

The mean number of training sessions (steps 1–6) for dogs to meet the acquisition criterion for all eight SPs was 39 sessions (SD=21) for Sequential training, 39 (SD=17) for Mixture training, and 39 (SD=16) for Inter-mixed training. All training sessions, including those in which dogs trained to (and failed to alert to) Vihta Vuori® 165 were included in this analysis. A linear model comparing the number of sessions to meet criterion predicted by training paradigm with a random intercept for institution showed no effect of training paradigm (X^2=0.03, df=3, p = 0.98).

The number of rounds of generalization tests required for dogs to reach criterion was also similar across groups. The Sequential group required an average of three rounds of generalization tests to meet criterion, the Inter-mixed group required an average of three and a half rounds, and the Mixture group required an average of four rounds to meet acquisition.

3.2. *Effect of SP odor on acquisition*

Little information is known as to whether certain smokeless powder varieties are more difficult to learn than others. To evaluate this, all training data were extracted from training steps 3–6 for all odors, encompassing about 12,500 trials. A generalized mixed effects model was fit to evaluate whether the probability of alerting was predicted by target odor with a random intercept for dog nested in institution. The model indicated a significant effect of target odor (X^2=372, df=7, $p < 0.001$). Estimated marginal means comparing performance across all target odors are shown in Fig. 4. The single base Vihta Vuori® 165 showed the lowest overall probability of response (12 of 18 dogs failed to meet criterion on this odor). The other single base (Accurate® 2015) and double base Vihta Vuori® 530 showed higher alert rates than Vihta Vuori® 165 but overall showed poorer performance during acquisition training compared to the remaining SPs, suggesting these are more difficult SP targets to learn and may require explicit training.

3.3. *Effect of training paradigm on generalization during first generalization test*

We compared the generalization rates to the four primary targets on the first generalization test across training groups. Our goal with this analysis was to determine if one training method led to higher learning rates for the primary odors after minimal training. Sequential, Inter-Mixed and Mixture paradigms all pose possible efficiency advantages. It is possible, as SPs are in the same explosive class and share common VOCs, that by training to one odor (Sequential training) we could achieve a high rate of generalization to other variants, making supplemental training to additional odors unnecessary. Training dogs to an odor mixture of multiple targets could also be an efficient training paradigm if dogs spontaneously generalized to the individual components within the mixture. We examined each training paradigms' probability to hold an alert to each of the four primary odors during the first generalization test (Fig. 5). We found that odor was a significant predictor of response across groups (X^2=60.0, df=3, $p < 0.001$). Vihta Vuori® 165 had the lowest probability of a response across all groups (see Fig. 5). This analysis indicates that regardless of initial training, dogs all struggled with Vihta Vuori® 165.

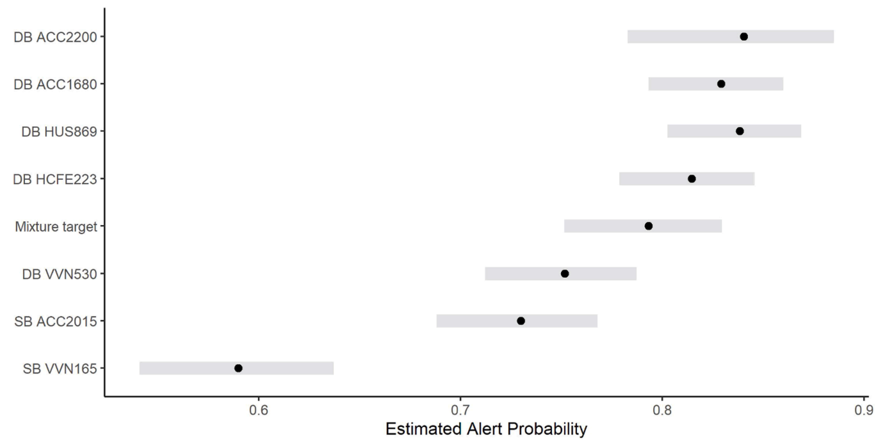

Fig. 4. Effect of SP Odor on Acquisition. Shows the estimated marginal mean and 95 % confidence interval for overall detection accuracy for each SP during training.

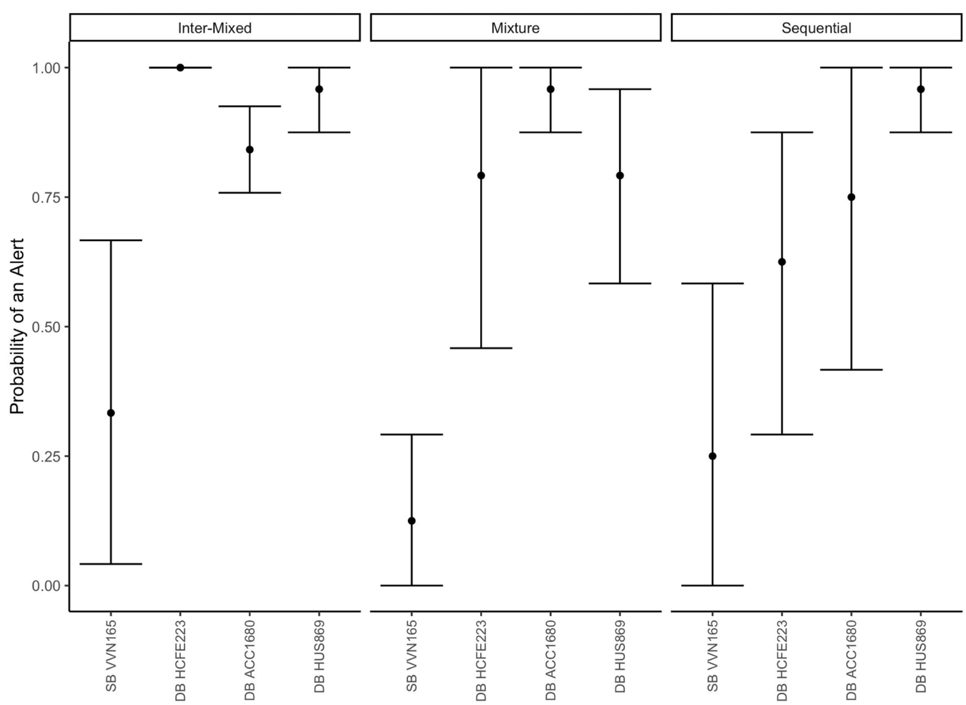

Fig. 5. Generalization during first generalization test to the four primary targets for the Mixture, Sequential and Inter-Mixed trained groups. SB = single base; DB= double base. VVN165 = Vihta Vuori® 165, HCFE 223 = Hodgdon® CFE 223, HUS869 = Hodgdon® US 869, ACC1680 = Accurate® 1680.

3.4. Effect of training paradigm on generalization during acquisition training

To evaluate whether a training paradigm led to higher rates of generalization during acquisition to the four variants (Accurate® 2015, Accurate® 2200, Vihta Vuori® 530, Hodgdon® H335) of the primary four targets, we combined the data across all generalization tests within the acquisition period and fit a logistic mixed effects model. There was no interaction between training paradigm and odor (X^2=9.58, df=6, p = 0.16). This term was removed to reduce model complexity. The reduced model indicated there was no statistically significant effect of training group on probability of an alert to a variant during generalization tests (X^2=1.98, df=2, p = 0.37). There was, however, an effect of odor (X^2=49.1, df=3, p < 0.001). Post hoc tests indicated that dogs were less likely to spontaneously generalize to Accurate® 2015 and Vihta Vuori® 530 compared to Accurate® 2200 and Hodgdon® H335, matching the overall poorer performance during acquisition for these two targets (see Fig. 6).

3.5. Effect of training paradigm on post-acquisition generalization

Following acquisition, all dogs completed a series of ten generalization sessions that were identical for all training paradigms. A logistic mixed effects model predicting probability of an alert as a function of training paradigm, odor type (single base or double base), and their interaction was fit with a nested random effect of dog in institution. The model showed no significant interaction (χ^2=3.76, df=2, p = 0.15). Therefore, the interaction term was removed to reduce model complexity. In the reduced model, there was an approaching-statistically significant effect of training paradigm on generalization to novel SPs (χ^2=5.75, df=2, p = 0.056, see Fig. 7) and no generalization difference between single base and double base powders (χ^2=0.13, df=1, p = 0.71).

Post hoc tests for effect of training paradigm indicated a trend that

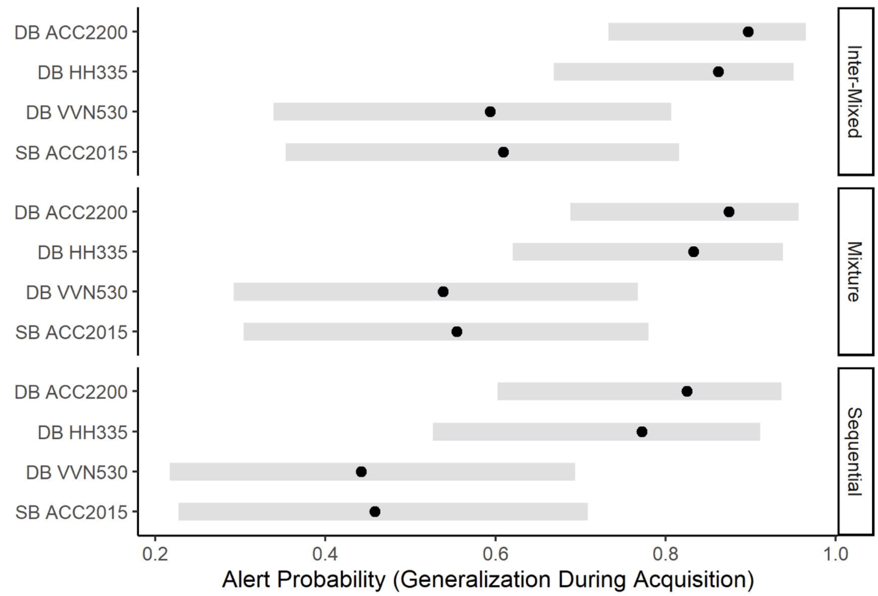

Fig. 6. Generalization during acquisition training. Shows the estimated marginal means and 95 % confidence intervals for the four odor variants tested for generalization during acquisition. SB = Single base, DB = double base, ACC2200 = Accurate® 2200, HH335 = Hodgdon® 335, VVN530 = Vihta Vuori® 530, ACC2015 = Accurate® 2015. Results are shown per training paradigm.

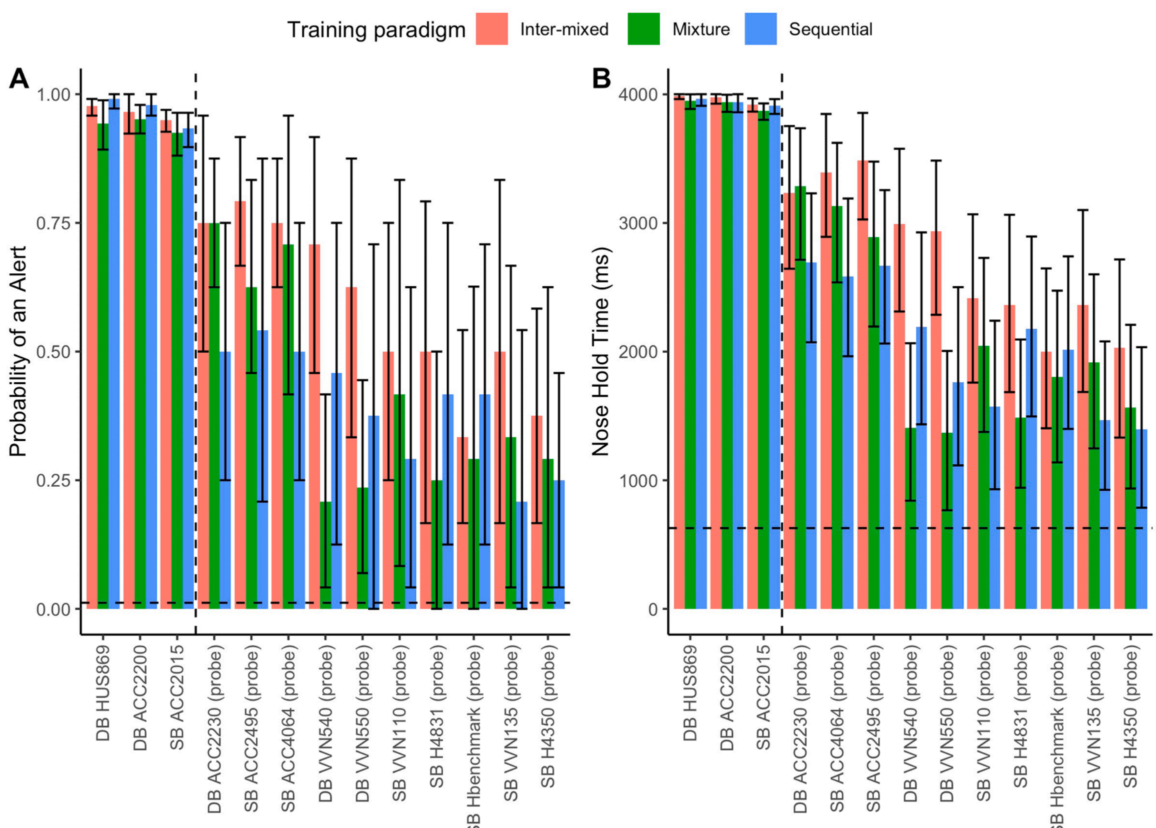

Fig. 7. Post-acquisition generalization. A: Shows the probability of an alert for the 10 post acquisition generalization odors. Dashed line shows mean alert rate to distractor odors. B: Shows the nose hold time for 10 post acquisition generalization odors. Dashed line shows average sniff time for distractor odors. Error bars who the bootstrap estimated 95 % confidence intervals. SB = single base; DB = double base, ACC = Accurate® brands, VVN = Vihta Vuori® brands, H = Hodgdon® brands.

Inter-mixed training led to higher overall probability to respond to the novel probes (67 %) compared to Mixture training (47.8 %, OR=2.26, z = 2.03, p = 0.06) and Sequential training (47 %, OR=2.36, z = 2.11, p = 0.06). There was no difference in generalization to novel SPs between Sequential and Mixture training (OR=1.05, z = 0.43, p = 0.91).

Total nose hold time showed a similar pattern, where there was no significant interaction between powder type (single or double base) and training paradigm (χ^2=2.72, df=2, p = 0.25) (see Fig. 7). A reduced model without the interaction indicated an overall effect of training paradigm on generalization to novel SPs (χ^2=6.02, df=2, p = 0.049) and no effect of SP type (single vs double; χ^2=0.04, df=1, p = 0.84). Post hoc tests indicate a trend that Inter-mixed training led to higher cumulative nose hold time to the novel probes (3.0 s) compared to Mixture training (2.3 s, t = 2.15, p = 0.08) and Sequential training (2.3 s, t = 2.17, p = 0.08). There was no difference in cumulative nose hold time to novel SPs between Sequential and Mixture training (t = 0.02, p = 0.92).

Fig. 8 shows the number of dogs that met proficiency for each SP odor (75 % or greater hit rate on probes) in the assessment for spontaneous generalization. More dogs met proficiency in the Inter-mixed group across most SP varieties, but most dogs did not meet proficiency for many SPs. This highlights that explicit training was still required for dogs to meet proficiency.

3.6. Effect of training paradigm on false alarm rate

To evaluate whether the training paradigm influenced overall false alarm rate during post-acquisition generalization tests, we fit a logistic mixed effects model in which probability of an alert to novel distractors was predicted by training paradigm. There was no effect of training paradigm on the probability of a false alarm (χ^2=1.33, df=2, p = 0.51), where the model predicted that the Inter-mixed group had a 4.6% false alert rate, the Mixture group had a 2.8% false alert rate and Sequential a had a 3.0 % false alert rate. The Sequentially trained dogs had a 15 % false alert rate to the charcoal pore strip; however this was the only distractor which had a false alert rate above 10 %, the recommended standard (see Fig. 9) (American Academy of Forensic Sciences, 2021).

Across all training groups dogs had a lower false alert rates on trials with a target present (Inter-Mixed = 1.7 %; Mixture = 2.2 %, Sequential = 2.3 %), compared to non-target trials (Inter-Mixed = 19 %, Mixture = 9.3 % and Sequential = 10.8 %)

3.7. VOC-based prediction of canine nose hold time

The VOC profile of each SP may yield important predictive value for dog generalization or overall performance for an SP variety. To explore the relationship between VOC profile and detection accuracy, we replotted the PCA of the VOC data used to determine SP variety selection for training. The same PCA is replotted in Fig. 10 with coloring reflective of mean detection accuracy for the odor. Trained varieties showed excellent detection, however, poor detection was observed for varieties that nearly overlapped with trained targets, indicating that PCA exploration of the physio-chemical space did not appear to overlap the canine psychological space (see Fig. 10).

As an alternative approach to PCA characterization of SPs, we used a Partial Least Squares regression (PLS) regression to evaluate whether VOC data were related to dog response in ways not captured by the PCA. To do this, we combined all training data from Steps 4,5,6 and generalization data during and post-acquisition. This yielded detection performance data from 9582 trials spanning 18 SP varieties and 18 dogs.

PLS regression was used to evaluate whether the peak area values from the 48 most frequently observed VOCs would yield a successful prediction of nose hold time. Cross validation identified eight components as the most efficient model. These eight components explained 73 % of the variance of VOC data and 17 % of the variance in nose hold time. Fig. 11 shows the assigned variable importance to each VOC for model prediction of nose hold time. TNT degradation products and 2,4-Di-tert-butylphenol were the most important compounds associated with canine performance. Of note, common SP signatures as analytical targets, such as Diphenylamine and Dibutyl phthalate, showed little

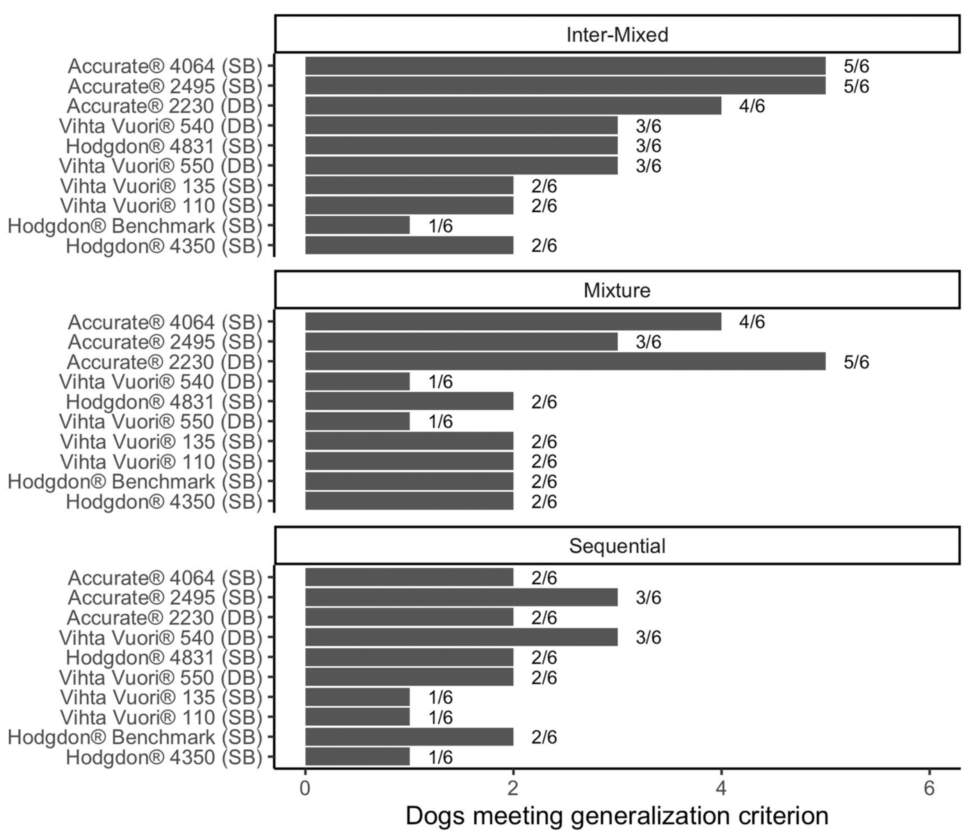

Fig. 8. Dogs that met the 75 % generalization criterion during the Post-Acquisition test in each group. Shows the number of dogs in each group that met generalization criterion.

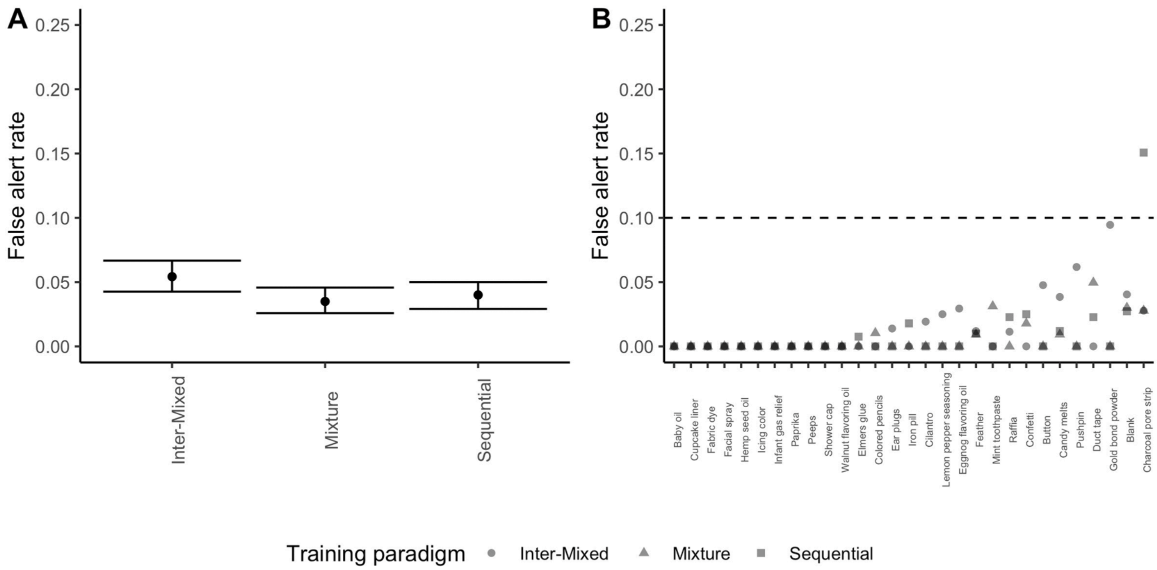

Fig. 9. Average false alert rates for each training paradigm, and average false alert rate by odor. Illustrates the false alert rate of each paradigm indicating to a non-target odor during the post-acquisition testing.

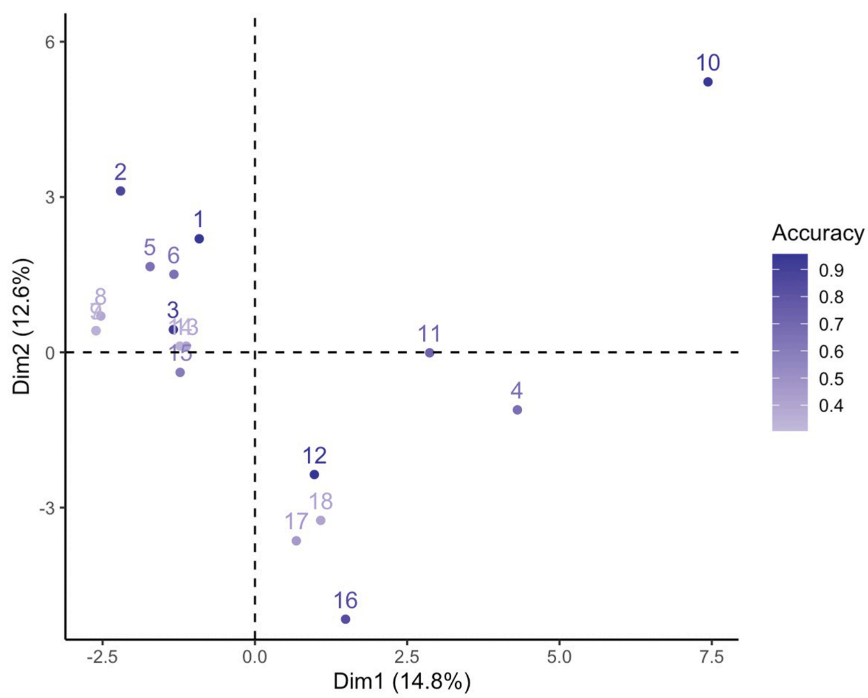

Fig. 10. Detection accuracy with SP PCA. Shows the PCA used to determine the SP varieties used with training. Color shows the dogs overall accuracy in detection of each target during training and post-acquisition generalization tests.

predictive relationship with canine nose-hold time. This is likely because these volatiles were present in all SP varieties (variants on which dogs generalized well, and those they did not), and thus were not predictive of nose hold duration.

Fig. 12 shows the relationship between nose hold time and relative VOC peak area for the top 10 most important VOCs identified by the PLS model. The relative VOC peak area was used to indicate the magnitude of that VOC observed amongst the SP varieties within the study. This figure shows that some molecules such as 2,4-Di-tert butylphenol were associated with low generalization rates whereas others were related to higher probability of response.

4. Discussion

All dogs reached acquisition criterion for the seven (if Vihta Vuori® 165 was dropped), or eight SPs (achieving >85 % detection accuracy and response during three out of four non-reinforced probe trials) within a similar time. Results suggest that enhanced generalization due to Inter-mixed training did not reduce training time. Interestingly, Mixture training had no effect on acquisition/training time, which is often the anecdotal rationale for its use, and was found to do so in previous research in rats (Keep et al., 2021). Mixture dogs did spontaneously generalize to three of the four components of their trained mixture, but they showed poor response to the fourth, Vihta Vuori® 165 (7 % alert

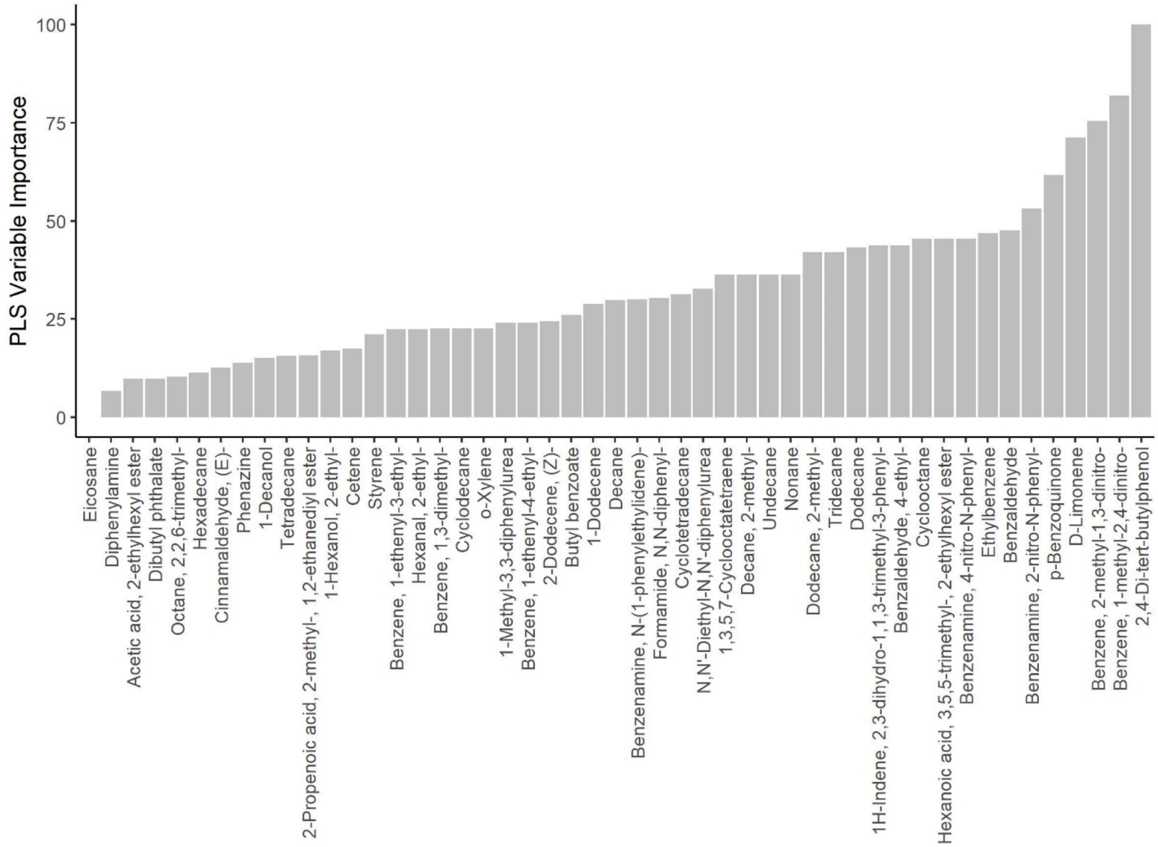

Fig. 11. PLS regression variable Importance. Shows the variable importance for each of the VOCs. Together, VOCs predicted 17 % of the variance in nose hold duration.

rate) (see Fig. 5). Even after responding to three of the four components of the mixture, they did not show enhanced generalization to the novel variants compared to the Sequential group, and thus overall training duration was not significantly different between these groups.

Issues detecting Vihta Vuori® were not unique to the Mixture group. Vihta Vuori® 165 was an anomaly. Twelve of eighteen dogs never met criterion on this target, whereas all dogs met criterion on all other training SPs. Anecdotal observation of dogs searching for Vihta Vuori® 165 suggested that Vihta Vuori® 165 was readily detectable through a change in behavior, but this target led to an aversive response that caused dogs to turn away and not complete the trained nose hold. Future work could explore canine detection of Vihta Vuori® 165 that allows dogs to make an alert away from the target, reducing any unpleasant experience with the odor that could influence response accuracy.

Results indicate modest effects of training paradigm on spontaneous generalization across novel SP exemplars, where there was a main effect of training group, and trend effects for the post hoc comparisons between Inter-mixed training compared to Mixture and Inter-mixed compared to Sequential training. This finding is in line with previous literature (Caldicott et al., 2024). Although generalization was enhanced, it remained below proficiency standards for detection of many SP varieties (i.e., 67 % alert rate for Inter-mixed vs 47 % for Mixture and Sequential). This result is consistent with findings that dogs trained to detect a double based smokeless powder were unable to generalize to a single base, or even to one of the primary VOC compounds in the headspace of SPs (Aviles Rosa et al., 2022). Our findings are also in line with previous literature, which shows that dogs fail to generalize to certain SPs without specific training (Oxley and Waggoner, 2009).

The underlying mechanism for the moderate improvement in generalization rates in the Inter-mixed group compared to the other two paradigms is uncertain. It is possible the improvement could be an effect of the Inter-mixed group receiving more reinforcement on targets (mean = 304 reinforced trials/target) compared to the other groups (Sequential mean = 277, Mixture mean = 281). We posit, however, that the slight increased reinforcement per target did not contribute to the improved generalization as the average numbers of reinforcements per target are similar across groups. This improvement could be due to continuous multiple exemplar training during Inter-mixed training. Multiple exemplar training has been shown to build a wider stimulus class and improve generalization to novel targets compared to rote learning of targets (Wright et al., 2017).

The overall poor generalization observed in our study suggests that Inter-mixed training may be an important part of a plan to ensure adequate generalization but is insufficient alone to generate the broad spontaneous generalization desired for an operational detection canine. These results suggest that to achieve 90 % or greater detection accuracy across novel SP varieties, more than the eight SP varieties used in training herein are likely needed. It remains possible, however, that by leveraging PLS regression from the data presented here, a smaller subset of SP varieties could be identified, leading to broader generalization.

Results from the PLS regression highlighted 2,4-Di-tert-butylphenol as one of the most important VOCs in nose hold behavior (see Fig. 12). This VOC was only present in the SPs Vihta Vuori® 165, Vihta Vuori® 135 and Vihta Vuori® 110. Dogs from all training paradigms had a mean alert rate < 50 % for these three targets, suggesting this VOC may be an important predictor of a non-alert response. This VOC may also be responsible for the anecdotally noted aversive response, given that Vihta Vuori® 165 has approximately 2x the peak area for this VOC compared to Vihta Vuori® 135 and Vihta Vuori® 110. This compound is a common compound found in essential oils and bacterial metabolites (Zhao et al., 2020).

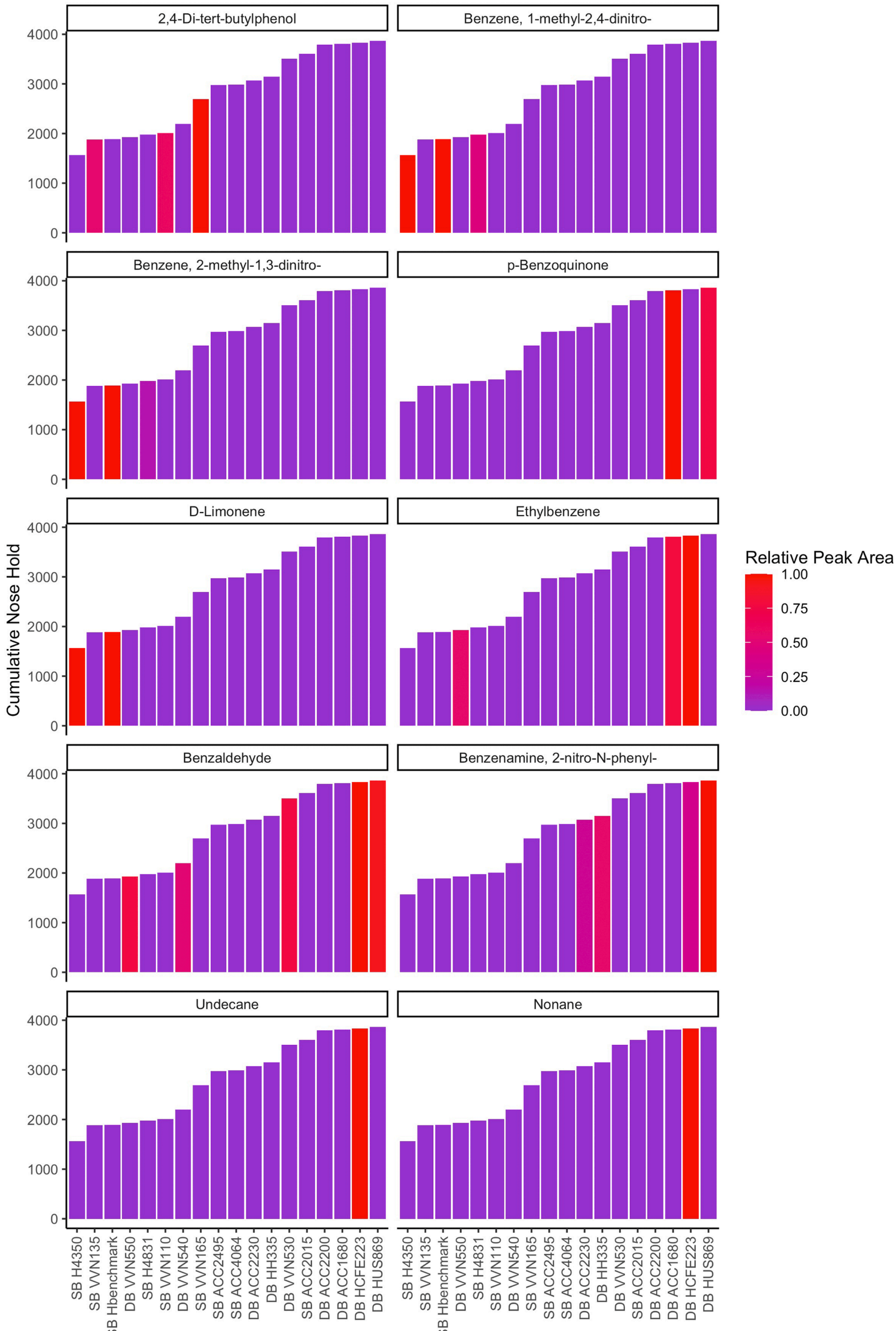

Fig. 12. PLS ten most important VOCs for smokeless powders. Each graph shows a different VOC in order of importance as ranked by the PLS model. Shows the total nose hold duration for each powder. Bar color shows the relative peak abundance that VOC was measure with respect to the 18 SPs evaluated. Thus, a one would indicate the powder showed the highest concentration for that VOC out of the 18 SPs measured. A value of zero would indicate that the VOC was not observed in that SP.

The second and third most important VOCs were Benzene, 1-methyl-2,4-dinitro- and Benzene, 1-methyl-1,3-dinitro-, which are likely TNT degradation products. These VOCs only appeared in Hodgdon® 4831, Hodgdon® 4350 and Hodgdon® Benchmark, which highlight the Hodgdon® single base product line, to which dogs showed uniformly poor generalization. Interestingly, single base versus double base was not an overall predictor of generalization, and these compounds were not observed in single base products to which dogs did show adequate generalization.

One important observation is that the most important compounds from the PLS showed that the addition of that compound caused generalization decreases. Previous research with rats has shown that contamination, or addition of a novel odor to a mixture, leads to greater discrimination compared to removal of one of the mixture components (Barnes et al., 2008). When selecting SP exemplars to train to it may be important to consider "off" odors in samples, not just odors that are unifying, such as 2-ethyl-1-hexanol, diphenylamine (DPA) and dibutyl phthalate.

The PLS regression results suggest that Vihta Vuori® and Hodgdon® single base products differ in important ways in the canine perceptual space from the trained Accurate® single base products that dogs met criterion for and showed adequate generalization. Our selection of SPs for training was based on PCA Component 1 and Component 2, which reflect the underlying variance in the VOC space of the powders. This, however, did not map clearly onto canine perceptual space (see Fig. 10). Based on the PLS regression, these results suggest that explicit training to Vihta Vuori® and Hodgdon® single bases may be critical to ensure adequate generalization across these product lines. Whether all single bases from these manufacturers need to be trained explicitly or if training with one product would be adequate, requires further testing.

Importantly, there were cases of excellent generalization observed during training. For example, all eighteen dogs spontaneously generalized to Hodgdon® H335 without any explicit training. Good generalization was also observed to Hodgdon® US 869 and Accurate® 2200, suggesting these may not require explicit training. Based on present results, a future study leveraging these canine generalization data may allow for development of an optimized series of SPs to maximize coverage of the canine perceptual space. At present, results suggest that less coverage may be needed across Hodgdon® and Accurate® double base powders and Accurate® single base lines, but more explicit training is necessary for Vihta Vuori® single bases, double bases and Hodgdon® single base lines.

There are important limitations to this work. First, sample size covered only six dogs per training paradigm. As always, extending these results to larger and broader populations of dogs would be important, especially given that the level of improved generalization from Inter-mixed training was useful, but modest. Second, our population of dogs had a variety of odor experiences from completely naïve, with all training occurring within the context of the experiment (TTU Cohort 2), to AU Cohort 1, which had years of prior odor work experience (although, importantly, not on SP or related odors). Given the small sample, we cannot determine the exact ramifications of these differences.

Another important limitation is the use of non-reinforced probes to assess generalization. Across the many testing sessions, dogs are repeatedly exposed to non-reinforced probes which could lower responding. We used intermittent schedules of reinforcement for the trained targets to help minimize the impact of four non-reinforced probes, but it does suggest that some of the lower generalization observed may reflect the repeated generalization probes. We selected this approach to ensure responses to probes reflected spontaneous generalization rather than learning due to reinforced responses to an odor. Dogs did continue to show responses to the various probes across the study period, suggesting that potential effects of the non-reinforced probes did not influence the groups differentially, therefore the relative comparison between training paradigms and SP varieties likely remain valid.

An additional limitation is that our random assignment of SPs within the PCA quadrants for a training odor and training odor variant over-represented double base powders in training. However, dogs were trained to two single bases, selection of odors was based on VOC profile, and results suggested single vs double base did not predict odor generalization. Nonetheless, it should be considered whether training to more single base varieties would have led to greater generalization. The present results suggest that two single bases were insufficient to get broad single base generalization.

Future studies in this area should explore other training paradigms for initial odor training, as well as best methods for maintenance training. There remain a range of potential variations of training paradigms using composite materials of SPs. Evaluating additional training paradigms would be an important next step for future studies. There is also a need to explore best practices for maintaining high rates of generalization during maintenance training once all exemplar odors are learned.

In conclusion, results suggest that Inter-mixed training led to moderately higher rates of generalization compared to Sequential and Mixture training, but that training paradigm did not change acquisition rate for SPs. These higher rates of generalization may yield significant operational improvements but did not lead to dogs reaching proficiency on novel targets (e.g. 90 % detection accuracy). This suggests explicit training to as many targets as possible within an explosives class, using an Inter-mixed method, will likely lead to the highest generalization rates. PLS regression with the VOC data highlighted that VOCs indicative of Vihta Vuori® and Hodgdon® single bases were different from the trained examples and led to poor generalization. Future work can leverage the canine behavioral data with PLS regression to create and evaluate an optimized list of SP varieties to maximize coverage for generalization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was funded by the Department of Homeland Security (DHS), Science and Technology Directorate, Detection Canine Program Office under contract #70RSAT22CB0000002. The material represents the position of the authors and not necessarily that of DHS. SAK's work was supported by the National Science Foundation Graduate Research Fellowship Program (DGE 2140745). Any opinions, findings, conclusions, or recommendations expressed in this work are those of the author (s) and do not necessarily reflect the views of the National Science Foundation.

We thank Chris Anzules for assistance with AU dog data collection and Holli Thompson, Ashley Ambery, and Nick Cassidy for care and management of the AU dogs. Additionally, we would also like to thank Avery Bramlett for her assistance with the TTU data collection, and the husbandry team at TTU for care and management of the dogs.

CRediT authorship contribution statement

S.A. Kane: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **G.N. Cupp:** Writing – review & editing, Writing – original draft, Investigation. **M. Rangel:** Writing – review & editing, Resources, Investigation. **A. Medrano:** Writing – review & editing, Resources, Investigation. **J. Davis-Miller:** Writing – review & editing, Investigation. **C. Collins-Pisano:** Writing – review & editing, Investigation. **B. Rogers:** Writing – review & editing, Resources, Methodology, Investigation. **D. Copeland:** Writing – review & editing, Resources, Methodology, Investigation. **L. Lazarowski:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **P. Waggoner:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **E.O. Aviles-Rosa:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **P.A. Prada-Tiedemann:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **N.J. Hall:** Writing – review & editing, Visualization, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.applanim.2025.106527.

References

American Academy of Forensic Sciences. (2021). *Standard for Training and Certification of Canine Detection of Explosives* (Standard 092).

Aviles Rosa, E.O., Fernandez, L.S., Collins-Pisano, C., Prada-Tiedemann, P.A., Hall, N.J., 2022. The use of an intermittent schedule of reinforcement to evaluate detection dogs' generalization from smokeless-powder. | SpringerLink https://link.springer.com/article/10.1007/s10071-022-01648-y.

Aviles-Rosa, E.O., McGuinness, G., Hall, N.J., 2021. Case study: an evaluation of detection dog generalization to a large quantity of an unknown explosive in the field. Animals 11 (5). https://doi.org/10.3390/ani11051341.

Barnes, D.C., Hofacer, R.D., Zaman, A.R., Rennaker, R.L., Wilson, D.A., 2008. Olfactory perceptual stability and discrimination. Nat. Neurosci. 11 (12), 1378–1380. https://doi.org/10.1038/nn.2217.

Bates, D., Machler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bureau of Alcohol, Tobacco, Fireams and Explosives. (2022). 2022 Explosives Incident Report (pp. 1–19).

Caldicott, L., Pike, T.W., Zulch, H.E., Ratcliffe, V.F., Wilkinson, A., 2024. The impact of training method on odor learning and generalization in dogs (Canis lupus familiaris). J. Comp. Psychol.

Fox, J., Weisberg, S., 2018. An R Companion to Applied Regression. SAGE Publications.

Furton, K.G., Myers, L.J., 2001. The scientific foundation and efficacy of the use of canines as chemical detectors for explosives1 Invited paper for the special issue of Talanta 'methods for explosive analysis and detection'.1. Talanta 54 (3), 487–500. https://doi.org/10.1016/S0039-9140(00)00546-4.

Gazit, I., Goldblatt, A., Grinstein, D., Terkel, J., 2021. Dogs can detect the individual odors in a mixture of explosives. Appl. Anim. Behav. Sci. 235, 105212. https://doi.org/10.1016/j.applanim.2020.105212.

Hall, N.J., Wynne, C.D.L., 2018. Odor mixture training enhances dogs' olfactory detection of home-made explosive precursors. Heliyon 4 (12), e00947. https://doi.org/10.1016/j.heliyon.2018.e00947.

Keep, B., Pike, T.W., Moszuti, S.A., Zulch, H.E., Ratcliffe, V.F., Porritt, F., Hobbs, E., Wilkinson, A., 2021. The impact of training method on odour learning and generalisation in detection animals. Appl. Anim. Behav. Sci. 236, 105266. https://doi.org/10.1016/j.applanim.2021.105266.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82, 1–26. https://doi.org/10.18637/jss.v082.i13.

Lazarowski, L., Dorman, D.C., 2014. Explosives detection by military working dogs: olfactory generalization from components to mixtures. Appl. Anim. Behav. Sci. 151, 84–93. https://doi.org/10.1016/j.applanim.2013.11.010.

Lennert, E., Bridge, C., 2018. Analysis and classification of smokeless powders by GC–MS and DART-TOFMS. Forensic Sci. Int. 292, 11–22. https://doi.org/10.1016/j.forsciint.2018.09.003.

Lenth, R. (2023). _emmeans: Estimated Marginal Means, aka Least-Squares Means_ [R package version 1.9.0]. https://CRAN.R-project.org/package=emmeans.

Oxley, J.C., Waggoner, L.P., 2009. Chapter 3—Detection of Explosives by Dogs. In: Marshall, M., Oxley, J.C. (Eds.), Aspects of Explosives Detection. Elsevier, pp. 27–40. https://doi.org/10.1016/B978-0-12-374533-0.00003-9.

Thomas-Danguin, T., Sinding, C., Romagny, S., El Mountassir, F., Atanasova, B., Le Berre, E., Le Bon, A.-M., Coureaud, G., 2014. The perception of odor objects in everyday life: a review on the processing of odor mixtures. Front. Psychol. 5 https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00504.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag https://ggplot2.tidyverse.org.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4 (43), 1686. https://doi.org/10.21105/joss.01686.

Wright, H.F., Wilkinson, A., Croxton, R.S., Graham, D.K., Harding, R.C., Hodkinson, H.L., Keep, B., Cracknell, N.R., Zulch, H.E., 2017. Animals can assign novel odours to a known category. Sci. Rep. 7 (1), 9019. https://doi.org/10.1038/s41598-017-09454-0.

Zhao, F., Wang, P., Lucardi, R.D., Su, Z., Li, S., 2020. Natural sources and bioactivities of 2, 4-di-tert-butylphenol and its analogs. Toxins 12 (1), 35.