Applied Animal Behaviour Science 287 (2025) 106638

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

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

Effects of training paradigm on dogs' (*Canis familiaris*) acquisition and generalization of odors when trained with multiple explosives classes

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

A R T I C L E I N F O

Keywords:
Dog
Olfaction
Explosives detection
Training
Generalization

A B S T R A C T

Explosive Detection Canines (EDC) are trained on a range of explosives and are expected to generalize to different and untrained explosive variants they encounter in field operations. This experiment investigated the effect of training method on detection dogs' rate of learning (acquisition) multiple target odors and their ability to generalize to variations of trained targets. Twenty-two dogs were randomly assigned to one of three training paradigms: Mixture, Sequential, and Inter-mixed. All dogs were trained to four explosive classes: (1) plastics/ RDX, (2) Det Cord/PETN-based, (3) AN (ammonium nitrate), and (4) TNT (trinitrotoluene). Dogs in the Mixture group were initially trained to a four-component mixture containing one variant from each of the four explosive classes. Sequential dogs were trained to one explosive variant at a time, reaching criterion on each odor before advancing. Inter-mixed dogs were trained to one variant from each class of explosives within the same training sessions. Dogs received generalization testing to another explosive variant in each class, and continued training until reaching criterion for eight variants. Post-acquisition generalization was tested to an additional 14 explosive variants. Overall, the training paradigm had no effect on acquisition time. Inter-mixed training led to higher rates of generalization to variants of TNT, plastics and PETN-based energetics tested post-acquisition compared to Sequential or Mixture training. Mixture and Sequential training did not differ in generalization rates. Although Inter-mixed training led to operationally relevant higher rates of generalization compared to the other two groups (e.g., 72 % vs 25 %-28 % for TNT varieties), it did not produce proficiency in detection for many novel variants from other classes such as PETN-based and plastic energetics. A partial least squares (PLS) regression model using VOC data to predict canine response time indicated that VOCs unique to novel variants were important predictors of low generalization rates of dogs. Inter-mixed training did lead to a slightly elevated false alarm rate (7 %) compared to Mixture (4 %) and Sequential (2 %) training, suggesting important considerations for the balance between generalization and false alarms operationally.

1. Introduction

Detection dogs (*Canis lupus familiaris*) are optimal biosensors because of their mobility, olfactory sensitivity in complex odor environments, and efficiency as a real-time threat detector (Furton and Myers, 2001). Dogs are deployed worldwide to locate targets including endangered species (Aviles-Rosa et al., 2023), infected human biological samples (Gokool et al., 2022), narcotics (Shellman Francis et al., 2019), and explosives (Hall and Wynne, 2018; Lazarowski et al., 2015; Lazarowski and Dorman, 2014). Despite their wide usage and efficacy as a biosensor, studies show that dogs struggle to generalize from trained targets to novel variants (Aviles-Rosa et al., 2022; Lazarowski and Dorman, 2014).

Generalization is a critical skill for explosives detection canines (EDCs) due to the diversity of their targets. There are many classes of explosives. Each class is organized based on the energetic material that dogs are trained to detect, and each has distinctive headspace odor compounds. While each explosive within a class share some common

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

https://doi.org/10.1016/j.applanim.2025.106638
Received 28 October 2024; Received in revised form 16 April 2025; Accepted 17 April 2025
Available online 18 April 2025

headspace volatiles, the headspace of explosives within a class can still vary enough to lead to within class generalization failures. For instance, prior research has found that dogs fail to generalize across different ammonium nitrate (AN) variants (DeGreeff and Peranich, 2021; Dorman et al., 2021). Similarly, Aviles-Rosa et al., (2022) showed that dogs trained to detect a smokeless powder did not generalize to one of the primary headspace compound characteristics of that class.

Research is needed to develop an EDC training procedure that will rapidly produce the highest degree of spontaneous generalization to explosive targets, while limiting false alerts to non-explosives. Currently most EDCs are trained using a sequential training method. In this training procedure, dogs learn to detect one target odor at a time until they reach proficiency and then they are trained on the next odor.

Another common training procedure is a “stew” or mixture method, in which dogs are trained to detect the mixed headspace of multiple explosive odors and are expected to generalize to each individual explosive within the mixture when presented individually. Mixture training could save training time by reducing the number of targets trained individually (Keep et al., 2021). Previous literature using mixture training for EDCs has shown some success (Gazit et al., 2021). Dogs trained to detect a mixture of three or five explosives did spontaneously generalize to the individual components of the trained mixture. Keep et al., (2021) compared olfactory training procedures in rats (including a mixture training method) and found that mixture training reduced training time but did not increase generalization to novel variations of the target over other training methods tested.

Mixture training relies on the assumption that dogs can perceive individual components as separate and identifiable odors within a mixture (elemental processing) (Moser et al., 2019). However, this assumption is not always true and one problem of using mixture training is an inability to predict how mixture effects impact the perception of the individual components. Overshadowing, when one salient component of the mixture is perceived more than others, is a potential mixture effect that could occur and limit conditioning to one or more of the explosives present in the ‘stew’ (Thomas-Danguin et al., 2014). Blending (or configural processing) is another mixture effect, in which all or some of the mixed odors combine to form a unique odor, which obscures perception of the individual components (Thomas-Danguin et al., 2014). In cases of overshadowing and blending, generalization to the individual components of the mixture could be reduced, thus making the mixture training procedure less advantageous for promoting generalization.

Keep et al., (2021) proposed a third odor training paradigm, Inter-mixed training. In this procedure, rats were trained to detect multiple target odors in the same session, but each odor was presented individually (not in a mixture). This is similar to some prior studies that used trial unique odor presentations to help facilitate generalization (e.g., Hall and Wynne, 2018). Keep et al. (2021) found that Inter-mixed training increased generalization to novel targets compared to both the mixture and sequential training. These results suggest that Inter-mixed training may be a beneficial training strategy for EDCs, who are required to generalize to untrained targets.

The goal of this study was to evaluate how Mixture, Sequential and Inter-mixed training paradigms affected the training duration required to learn eight odors from four explosives classes (plastics [RDX-based], detonation cords [PETN-based], nitrate fertilizers, and TNT varieties), and generalization to novel variants of explosives within each class. This work was part of a larger study that explored the effects of training paradigm when training multiple explosive classes (this study), and when training multiple varieties within the same explosive class (within explosives class; (Kane et al., 2025)).

2. Methods

2.1. Participants

A total of 22 dogs from Texas Tech University (TTU) and Auburn University (AU) participated in this experiment. The dogs completed the study in cohorts of six, with two cohorts from each university. TTU acquired and trained 12 naïve dogs from a local animal shelter as part of the Training for Adoption program. All dogs from TTU were mixed breed dogs and were spayed/neutered. Work at TTU was reviewed and approved for ethics under TTU IACUC protocol number: 2022–1180.

AU initially selected 13 dogs for this experiment. Dogs were purpose-bred detection dogs from the AU Canine Performance Sciences detection dog breeding and training program and had varying amounts of odor experience. However, all dogs were naïve to the explosive odors used in this experiment. Three of the initial thirteen dogs selected were removed from the study due to motivation issues. All dogs from AU were Labrador retriever and Labrador retriever crossbreeds, and all but one female were spayed/neutered. AU dog activity was reviewed and approved by the AU IACUC (protocol number: 2022–5029). See Supplemental Table 5 for demongraphic information.

One of the four cohorts participated in the related within explosives class study prior to the current study and thus had prior training with smokeless powders (Kane et al., 2025). Dogs that completed both studies did not change training paradigms between studies due to the potential for training paradigm effects to be maintained across time and different odors. Thus, to maintain potential effects of the training paradigm, rather than potentially confound them with historical effects from the other study, the training paradigm was kept consistent across the two experiments. For example, a dog trained using the Mixture paradigm in the within class study was also trained in a Mixture paradigm in this study. All dogs had stratified random assignment to each cohort.

2.2. Experimental design

Dogs were trained using one of three training paradigms (see Fig. 1) in an automated three-choice olfactometer (described in detail below). This allowed for training and testing to be conducted double-blind. During acquisition, dogs were trained to detect eight different explosives from four classes: RDX/plastics based, TNT-based, AN variants, PETN-based (2 per explosive class). For each explosive class, one odorant served as a primary target and the other served as a variant to measure generalization during acquisition (see Odorants). The odors were the same for every training group, and every cohort. Broader generalization to 14 novel explosive varieties from the same four classes was assessed in a post-acquisition test (see Odorants and Table 2 for details).

2.2.1. Training paradigms

2.2.1.1. Sequential training. In this paradigm dogs were trained to detect one explosive at a time. Once dogs met criterion on the first odor (see Training Procedures below for more details on training criteria), generalization was assessed to all eight targets. Dogs were then trained to the next odor which they failed to generalize to on this test (see Table 2 for odor orders). This cycle of training, testing, training, continued until dogs had acquired all eight targets.

2.2.1.2. Mixture training. Dogs were trained initially to detect the mixed headspaces of the four primary targets, one from each class. Once meeting criterion on this mixture, generalization was tested to each of the 8 training odors. Dogs were then trained in a sequential manner until reaching criterion on all eight odors to match standard practices with odor mixture training.

2.2.1.3. Inter-mixed training. Dogs in this paradigm were trained to detect four primary targets (one from each explosive class), where each odor was presented every training session, but presented in individual trials. Once dogs met criterion on the primary four odors, generalization was assessed to all eight odors. Dogs were trained to odors which they

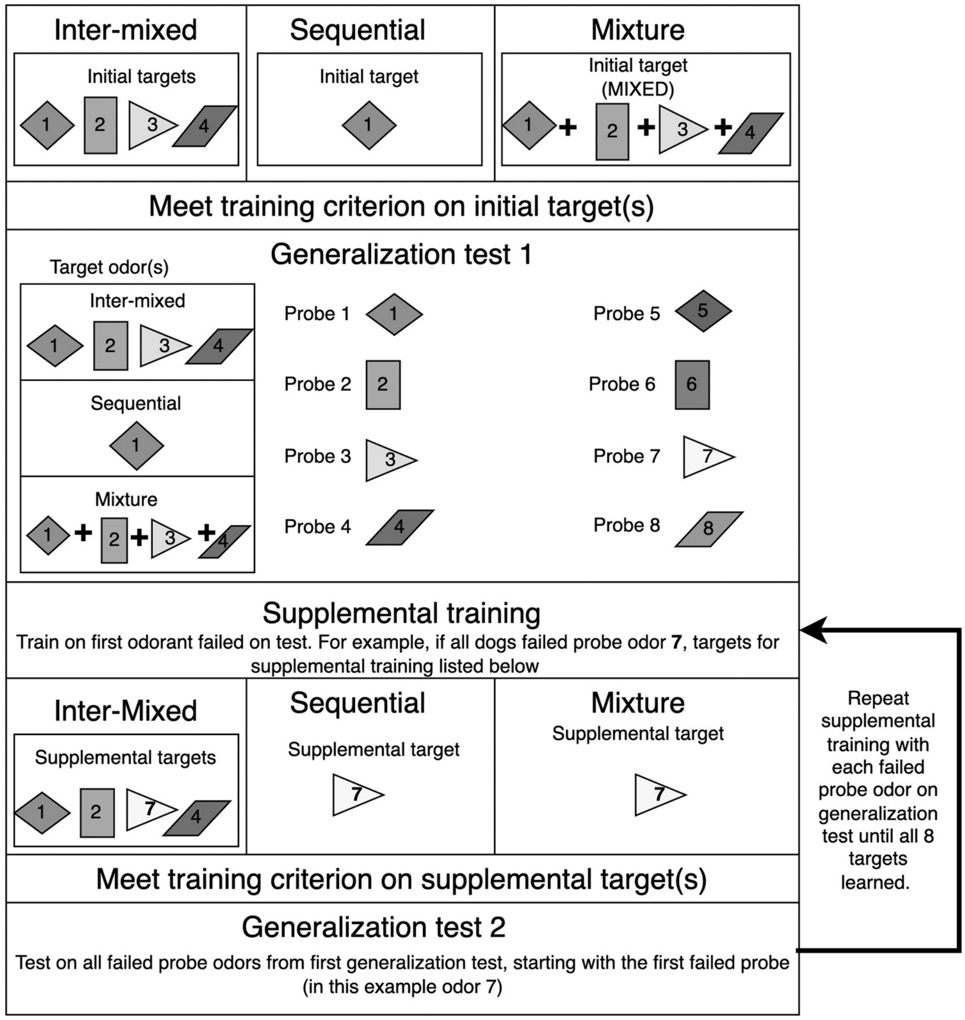

Fig. 1. Experimental design for the three training paradigms. Shapes represent odors in the same explosive class. Each shape has a distinct hue to indicate that odors in the same class may not share the exact same odor signature. Supplemental training occurs for the first odor failed in the dog's assigned sequence. Training and generalization testing cycle until dogs meet criterion on all eight training odors. For the inter-mixed group, the supplemental trained odor substitutes for the related odor class (e.g. a TNT variant for TNT).

failed to generalize to in an Inter-mixed method. The failed odor replaced the primary target of the same explosive class. Dogs were trained until they met criterion on all eight targets.

2.3. Equipment

Three independent, identical automated odor delivery devices (olfactometers), each with the capacity to hold 12 odor samples, were used in this experiment. Each olfactometer was controlled via Bluetooth by a computer program. These olfactometers provided a fast odor presentation system with a high degree of stimulus control. Briefly, each box functioned by pumping two liters of air through a charcoal filter to remove contaminants. The air was split into two lines, each controlled by a rotameter. One rotameter controlled the dilution line, or the clean air that mixed with the odor headspace. The other rotameter regulated air flow to the sample through a set of solenoid valves. These valves were controlled by the computer program, and if open, air moved through these valves to another rotameter. This second rotameter allowed for control of air flow over the sample to create odor mixtures. As air was pushed into the sample vial, an equal volume of air was displaced from the vial and flowed through a one-way check valve into a Teflon manifold. This one-way check valve prevented back flow and thus contamination between samples. The odor from the sample and the diluting clean air were mixed in the mixing manifold. Air flowed from this mixing manifold to an odor sampling port.

For sequential and inter-mixed training, all odors were presented at 1:1 dilution with clean air (i.e., 50 % dilution) at a total flow rate of 2 standard liters per minute (SLPM). During mixture training, the target mixture was presented with no air dilution, producing a total flow of 2 SLPM comprised of the four odors contributing an equal 0.5 SLPM to the final mixed odor (i.e., all equal parts). The distractor (non-target) odors for Mixture training were diluted in a 1:1 ratio.

Air flowed from the mixing manifold into an odor sample port, in which dogs were trained to put their nose. Each odor port had an IR beam pair that measured the time the dog's snout was inserted in 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. Handlers (via a Treat & Train® at AU, and via hand at TTU) gave the dogs food as a reinforcer for a correct response. If the dog did not correctly hold an alert (no nose hold for required alert duration) or alerted to an incorrect olfactometer a different tone would play, indicating to the dog and the handler that the dog was incorrect and to not provide a reinforcer. In addition to these two sounds, there was a trial start tone and a trial end tone. For a more detailed description of the olfactometers used in this study, please see (Aviles-Rosa, et al., 2021).

The olfactometer was loaded to include the four primary targets, a spot for an odor variant on probe trials, at least one empty vial, and five distractor odors. Distractor odors were changed approximately every week to ensure dogs were not learning to avoid distractors, rather than to learn a target odor. During every generalization test two novel distractor odors were present.

2.4. Initial olfactometer training

All dogs received training to use the olfactometer prior to introducing the explosive odors. Initially TTU canines were trained to search the boxes for food odors (hot dogs), as these dogs had no known prior search experience. Once these dogs showed interest in the olfactometers and held at least a 1 s nose hold in the box with the target odor, they moved on to searching for non-food odors. All dogs were trained until they reached 90 % accuracy at detecting a 1 % solution of 1-bromooctane and a 1 % solution of amyl acetate. During initial training dogs learned to search the olfactometers without handler intervention, maintain a 4-s nose hold in the target olfactometer and correctly clear all olfactometers non-target trials was present ("all-clear"). During a correct all clear response, the dog would put their nose into each olfactometer (as recorded by the IR beam pair) and leave the search area without holding an alert to any of the olfactometers. After ten seconds in which the dog did not break an IR beam pair (i.e., did not check an olfactometer) a correct tone would play. No explicit behavior was trained to indicate an all-clear response. Dogs were also trained to work under an intermittent schedule of reinforcement where 40 % of trials were non-reinforced no matter if the response was correct or incorrect. During non-reinforced trials the trial end tone would play after a response (a nose hold in an olfactometer), or if the dog checked all olfactometers and did not continue break an IR beam for ten seconds ("all clear" response). Throughout this study dogs were allowed to search ports multiple times. This training was done to ensure that dogs were familiar and comfortable with the experimental apparatus.

2.5. Training procedure

The training procedure and generalization testing protocols for each method were held as identical as possible to standardize the number of sessions needed to reach criterion. All training sessions were comprised of 20 trials. See Table 1 for training step details.

In the first stage (step 1) of training, dogs completed two sessions of 20 trials, with a nose hold time on the target of 0.1 s. The goal of this stage was to reward the dog for putting their nose in the odor sampling port with the target. During this initial step there were no programmed consequences to the dogs for holding their nose and smelling other odors for longer than 0.1 s. Thus, the program "waited for the correct" (WFC) response. During step 1, the odor was present in one of the three olfactometers every trial (there were no non-target trials), and the handler was unblinded and aware of the port containing the target.

Once dogs completed the initial two training sessions, they automatically progressed to step 2. In step 2 the nose hold time for an alert (both correct and incorrect) was increased to 1 s while all other parameters remained unchanged. Dogs were required to achieve 90 % accuracy (18 out of 20) on step 2 to proceed to the next step.

Because some of the targets were more challenging to learn, additional procedures were added if a dog failed to meet criterion after multiple training sessions. If, after five sessions on step 2, a dog did not achieve 90 % accuracy, the dog repeated two sessions at step 1. The dogs then moved back to step 2. If, after another three sessions at step 2, the dog still did not achieve the 90 % criterion, they would go back to step 1 for a final time, complete two sessions and move back to step 2. If after another three sessions at step 2, the dog did not achieve 90 % accuracy, they were advanced to step 3. Dogs were advanced to the next step because modifications such as increased alert time and introduction of blanks could help facilitate discrimination for more difficult targets and because continuing training at the same step was unlikely to improve performance.

During step 3, non-target (blank) trials were introduced in four of the 20 trials. On these trials, the olfactometers would not present the target odor (only distractors) and dogs were required to search each box and not alert to any box. These correct "all clear" responses were reinforced the same as correct responses to the target odor. Further, the alert time was modified to be variable between 0.1 s and 4 s to train increased alert times. Lastly, the handler was blinded to odor location. Dogs were required to make no more than two incorrect responses on trials in which the alert time was greater than 0.5 s to move to the next step. Trials with alert times less than 0.5 s were not counted toward overall accuracy. The program was set to "wait for correct" on this step. If after 11 sessions at step 3, dogs had not met criterion (90 % correct responses), dogs were automatically moved to step 4. This was done to allow for increases in alert time and incorrect responses to terminate a trial to facilitate discrimination by increasing the response cost.

On step 4 the alert time was changed to 4 s, and trials terminated if the dog made an incorrect response (i.e., the program did not wait for a correct response). All other settings were consistent with step 3. Dogs were required to achieve 90 % accuracy to move onto the next step.

Steps 5 and 6 introduced non-reinforced trials where the trial ended with an end of trial tone regardless of the response of the dog (a correct or incorrect response). These "ambiguous" trials were used to prepare dogs for non-reinforced probe trials during generalization testing. In step 5, 20 % of trials were non-reinforced trials. Dogs were required to complete at least two sessions at step 5, achieving an accuracy of 85 % or greater on two consecutive sessions. For step 6, the non-reinforcement rate was increased to 30 %. Once dogs hit the minimum criterion of 85 % accuracy on two consecutive sessions of step 6, they moved onto generalization testing.

2.5.1. Inter-mixed training target weight adjustments

All dogs in the Inter-mixed training group started training with four targets weighted so each would appear in blocks of three consecutive trials. The order in which these blocks were presented for each target was randomized within a session (i.e., the order of targets). If a dog required five or more sessions at step 2 or any other step, the target odors were weighted to increase presentation frequency of odors on which dogs performed poorly. Target weighting was done proportionally so that the dogs' overall accuracy on each target determined the frequency each odor appeared. This procedure weighted the presentation of odors

Table 1
Training procedure for experimental training.

Step	Alert Time (s)	Wait for Correct	Blinded	Blanks (% of trials)	Intermittent Reinforcement (% of reinforced trials)	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)

Table 2
Explosive target odors. Includes types of explosives trained, or probed in each explosive class, and the order each cohort was presented the odors. Note that Inter-mixed group dogs were trained to the four primary targets at the same time, and the Mixture group dogs were trained to the headspace of the four primary targets initially.

Explosive class	Type	Use	Order AU cohort 1	Order TTU cohort 1	Order AU cohort 2	Order TTU cohort 2
Plastics	C4 (M112)	Primary target	1	2	4	3
Plastics	PE−4 (untagged)	Variant target	5	6	8	7
Nitrate fertilizers	Fertilizer grade AN	Primary target	3	1	2	4
Nitrate fertilizers	Calcium AN	Variant target	7	5	6	8
Trinitrotoluene	American flaked TNT	Primary target	4	3	1	2
Trinitrotoluene	French flaked TNT	Variant target	8	7	5	6
PETN-based	Detonation cord 100 g/ft	Primary target	2	4	3	1
PETN-based	C2 Sheet explosive	Variant target	6	8	7	5
Plastics	C4 (M112) (untagged)	Post-acquisition Probe	4	5	11	10
Plastics	Semtex	Post-acquisition Probe	5	4	9	11
Plastics	C4 (Tripwire)	Post-acquisition Probe	1	8	14	7
Plastics	PW4	Post-acquisition Probe	6	3	10	12
Plastics	Bofors C4 (untagged)	Post-acquisition Probe	2	7	12	8
Plastics	Bofors C4	Post-acquisition Probe	3	6	13	9
PETN-based	Detonation cord (400 g/ft)	Post-acquisition Probe	8	13	6	2
PETN-based	Detonation cord (10 g/ft)	Post-acquisition Probe	9	12	8	3
PETN-based	Flex-X	Post-acquisition Probe	7	14	7	1
Trinitrotoluene	Canadian TNT	Post-acquisition Probe	14	9	3	6
Trinitrotoluene	Cast (powdered)	Post-acquisition Probe	12	10	2	5
Trinitrotoluene	Cast	Post-acquisition Probe	13	11	1	4
Nitrate fertilizers	Large prill	Post-acquisition Probe	11	1	5	14
Nitrate fertilizers	AN (kinepack)	Post-acquisition Probe	10	2	4	13

on which dogs performed more poorly to occur more frequently and decrease the frequency of odors on which dogs performed well (i.e., an odor with a weight of two would appear twice as often as a target with a weight of one). To weight the odors, we subtracted the dog's accuracy for an odor from 100 %. We multiplied this by the maximum weighting of three (three consecutive trials of the same odor). We rounded the resulting product up to the nearest whole number. For example, if a dog achieved 90 % accuracy for a specific odor, we multiplied (1–0.9) * 3 = 0.3 and rounded to the next highest number of one. If a dog achieved 50 % accuracy (1–0.5) * 3 = 1.5, the weight would be 2. If a dog had unequal weights after meeting criterion on Step 3, the odors were re-weighted, so they all appeared at an equal frequency (1:1:1:1).

2.5.2. Generalization Testing During Acquisition

The purpose of testing dogs' generalization throughout acquisition was to determine both if the dogs had learned the target odor(s) in non-reinforced probe trials, and whether they would spontaneously generalize to other untrained explosive odors, negating the need for explicit training to that target. Dogs were tested on all eight targets during their initial generalization test. Two novel distractor odors were added for each generalization test, to ensure that alerts to the probe odors were not simply due to novelty.

Generalization testing consisted of 20 trials, two of which were non-reinforced presentations of the trained target or non-target trials, four of which were non-reinforced probes of the variant being tested, and 13 reinforced trials. The probe trials evaluated responses to the odor without explicit reinforcement. Following any response (alert or no alert) probe trials terminated with the "end of trial" tone without the "correct" or "incorrect" tone or reinforcer. The overall number of non-reinforced trials during generalization testing was consistent with the reinforcement rate of the final training session (step 6).

An odor was considered successfully trained if the dog correctly alerted to that odor on at least three of the four probe trials in a generalization test session. That odor was then removed from further training and testing for that dog. If a dog did not alert on at least three of four probes, that odor was considered insufficiently trained. Dogs received "supplemental training" (described below) for these targets. In cases where a dog previously received explicit training and met training criterion for an odor but failed to reach criterion (3/4 probes) during a generalization test, the dog started supplemental training at step 6, to re-meet step 6 criterion for that odor. If the odor was not previously trained explicitly (e.g., a target variant or untrained primary target), supplemental training started at step 1. If a mixture-trained dog failed to meet this probe trial criterion they moved to supplemental training starting 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 order assigned to that cohort (see Table 2 for order). Upon meeting criterion of the new odor, generalization testing was repeated for all odors for which dogs did not previously meet criterion. The supplemental training followed by the generalization test cycle was repeated until dogs met criterion on all eight target odors. This train-test-train-test cycle was repeated to identify the precise time dogs would meet criterion on all eight odors, allowing the possibility that dogs may not require explicit training to some varieties.

2.5.3. Supplemental training

Supplemental training was used to train dogs to odors which they did not spontaneously generalize to during a generalization test. This followed a similar progression through the six training 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 Sequential training. Thus, these two groups only differed in the initial training to a single odor or the 4-component mixed odor.

The Inter-mixed training group followed an inter-mixed supplemental training by having the first failed generalization odor replace one of the four primary training odors. The new odor was weighted to appear more frequently (3x as frequent as the others) to facilitate training to the new odor.

2.6. Post-acquisition generalization

After dogs showed proficiency in detecting the eight explosive varieties during acquisition training and testing, generalization was tested with 14 additional explosive varieties (see Table 3). The goal of this experiment was to determine if one training paradigm led to greater generalization to novel variations of the trained explosive classes.

2.6.1. Post-acquisition generalization: training

Dogs received refresher training for their initial experimental paradigm at step 4 settings (Mixture received the mixed odor, Sequential their first odor and Inter-mixed the four primary targets) immediately before testing began. Dogs were required to achieve an average score of 90 % or greater across two 20-trial sessions on their original training paradigm.

2.6.2. Post-acquisition generalization: testing

The generalization test was identical for all dogs across experimental training groups, ensuring that only prior training history influenced generalization probability. For each session, a trained explosive target that matched the class of the variant being tested served as the target odor. The target variants of the same explosive class (e.g., nitrate fertilizers, plastic explosives, PETN-based, or TNT based) were then assessed for that target.

Each day of generalization tests started with an initial refresher session to ensure that performance to the primary target was satisfactory. If a dog scored greater than 80 % across ten trials, the dog advanced to probe testing for the day. Any score lower than 80 % resulted in a re-test of another ten trials until they reached criterion.

Each generalization test consisted of one target explosive, four novel distractors, and one novel variant probe for that target explosive class. The four distractors were novel at the start of every generalization testing day. Two to three generalization tests occurred per day. Of the 20 trials in each testing session, four were non-reinforced probe trials, one was a non-reinforced non-target trial, two were non-reinforced target trials, and 12 were reinforced target and non-target trials.

2.7. Odorants

The four primary targets included: fertilizer grade ammonium nitrate (AN), American flaked Trinitrotoluene (TNT), detonation cord (Det Cord: 100 g/ft), and Composition4 (C4). The four variants of the targets were: Calcium Ammonium Nitrate (CAN26), French Flaked TNT (TNT FF), Pentaerythritol tetranitrate (PETN) Sheet Explosive, and Untagged Pe-4, respectively. Target materials were sourced from Auburn University. The use, handling, and storage of all explosive materials was approved by the respective institutional Laboratory Safety Committee. All TNT and PETN-based explosives were untagged. All plastic explosives were tagged unless otherwise noted in the figures and tables.

The order of training for the targets was assigned per cohort (see Table 2). This order was balanced across cohorts such that each explosive type was trained in each position (1st, 2nd, 3rd, 4th). The order for variants (odors five through eight) followed the order for the related primary target.

An additional 14 target variants (2–6 for each class) were used to assess generalization on the post-acquisition generalization task. This group of novel targets included six plastics, three PETN-based, three TNT variants, and two nitrate fertilizers (see Table 2). Four dogs (2 from sequential, 1 mixture, 1 inter-mixed) did not receive testing with the TNT variants but did test with all other variants.

At AU explosives were additionally inside a secondary glass container and two anti-static bags. At TTU explosives were stored in sealed glass containers with a Teflon/silicone lid, with a secondary wrap of parafilm. Both universities stored their explosive samples in a bunker.

Additional information on the chemical characterization of the explosives is available in supplemental materials.

2.8. Statistical analysis

All statistical analyses were conducted in R version 4.3.2. Data was initially cleaned using dplyr and tidyverse packages (Wickham et al., 2019). Plots were created using ggplot2 (Wickham, 2016). Analysis of variance (ANOVA) analyses were done using the car (Fox and Weisberg, 2019), lme4 (Douglas Bates et al., 2015), and lmerTest (Kuznetsova et al., 2017) packages. We conducted post-hoc analyses through the emmeans package using the false discovery rate post-hoc correction for all contrasts (Lenth, 2024).

For all models, we used a mixed-effects approach. To account for potential differences between locations (Institution: TTU or AU), we fit a random effects structure in which the dog was nested within their respective institution. To evaluate the effect of training paradigm on probability of an alert, a logistic mixed effect model was used to account for the binomial (correct vs incorrect) outcome. For the nose hold time, linear mixed effect models were used. To prevent incidental re-setting of the IR beam break timer artificially increasing nose hold time, the maximum nose hold time was set to 4 s. The maximum 4 s nose hold duration was exceeded on some trials by some dogs due to their nose not effectively breaking the IR beam (due to a narrow muzzle). In addition, all 0 s nose holds were excluded for all analyses because this indicated that the dog did not sample that olfactometer and was therefore not a measure of response to the odor.

To evaluate whether the chemical characterization predicted generalization, we conducted a Partial Least Squares regression for each class of explosive. The peak area for each VOC identified in the chemical characterization was used as a predictor in the PLS. The outcome measure was the dog's nose hold time for that odor during post-acquisition generalization testing. We only interpreted models for an explosive class that explained greater than 5 % of the variance in nose hold time. To enhance visualization of results for the 10 most important VOC predictors, we calculated relative peak area abundance by dividing peak area of a VOC for an explosive by the maximum peak area observed for that VOC across the sample set.

3. Results

3.1. Effect of training paradigm on acquisition

The mean number of training sessions for dogs to meet the acquisition criterion for all eight odors was 87 sessions (SD=24) for the Sequential group, 86 (SD=23.8) for the Mixture group, and 67 (SD=15) for the Inter-mixed group. A linear model was used to compare the number of sessions to meet criterion predicted by training paradigm with a random intercept for institution. While there was a statistically significant effect of training paradigm (χ^2=6.40, df=2, $p = 0.04$; Fig. 2A); the false discovery rate adjusted post hoc tests indicated only a trend in which Inter-mixed dogs met criterion 20 sessions faster than Sequential ($t = 2.35$, df=18, $p = 0.07$), but no other comparisons neared significance ($p > 0.17$). This only trend-level post-hoc difference appears driven by a lack of duration variety observed in the more experienced dogs at AU. These dogs showed similar numbers of sessions across all groups (68 sessions for Inter-mixed, 68 for Mixture, and 66 for Sequential). Naïve dogs at TTU, however, showed that Inter-mixed dogs met criterion within mean of 66 sessions (like AU), but Mixture dogs (mean of 98 sessions) and Sequential (mean 107 sessions) required more sessions (see Fig. 2B).

3.2. Effect of mixture training on generalization to individual components of the mixture

Immediately following meeting criterion with Mixture training with the TNT, C4, Detonation Cord, and AN mixture, dogs completed a series of generalization tests to each component. Fig. 3 shows generalization to each odor of the mixture when presented separately (during the first generalization test, prior to any supplemental training). We fit a generalized logistic mixed effect model, which predicted probability of an alert by odor, with a random effect of dog nested in institution; however, this model showed complete separation and failed to fit. We subsequently fit a linear mixed model with probability of alert predicted by odor with a random effect of dog nested within institution. This model indicated that dogs did not generalize to all explosives in the mixture equally (χ^2=45.6, df=3, $p < 0.001$). A post-hoc analysis showed higher generalization rates for C4 compared to AN ($t = -5.30$,

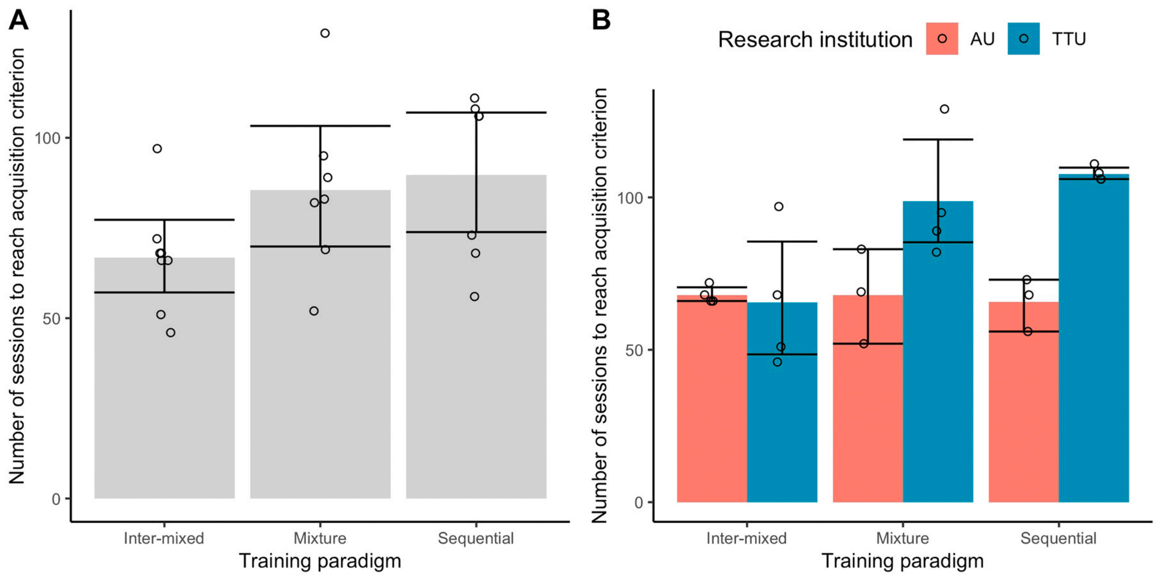

Fig. 2. Effect of training paradigm on acquisition. A: Shows the relationship between number of sessions for each dog to reach acquisition criterion by training paradigm. Individual data points are shown for each dog. B: Illustrates the relationship between institution and training paradigm. AU = Auburn University; TTU = Texas Tech University. Error bars show bootstrap estimated 95 % confidence intervals.

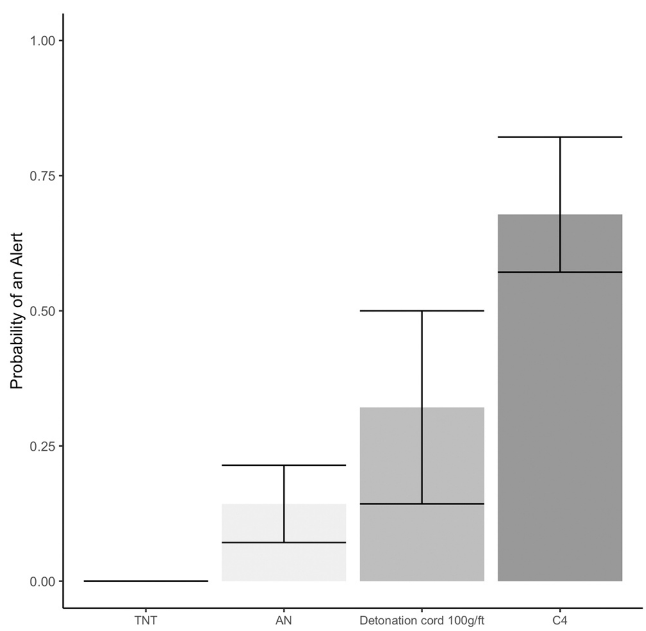

Fig. 3. Effect of Mixture Training on generalization to components of mixture on the first generalization test. Shows the probability of a mixture trained dog responding to the components in the mixture on the first generalization test during acquisition training. Error bars show bootstrap estimated 95 % confidence intervals.

$p < 0.001$), TNT ($t = 5.77$, $p < 0.001$), and detonation cord ($t = 3.54$, $p = 0.013$). The only other contrast that was approaching significance was between Detonation Cord and TNT ($t = 2.74$, $p = 0.061$). All other contrasts were non-significant, showing that dogs largely only alerted to C4 with little to no response to the remaining components of the mixture.

3.3. Explosive odor acquisition

Although dogs are widely trained to a range of energetic classes, there is little information on whether some odors are more challenging to learn than others. We extracted all steps 3–6 training data for each odor across the acquisition phase (the phase in which the dogs were explicitly trained to detect the four primary and four variant odors) for all odors encompassing about 30,000 trials. This analysis aimed to evaluate "training difficulty", by measuring accuracy across the reinforced training period. A generalized mixed effect model was fit to evaluate whether the estimated proportion correct was predicted by target odor with a random intercept for dog and institution. The model indicated a significant effect of target odor, indicating some odors were more difficult to learn (χ^2=556, df=8, $p < 0.001$). Post hoc tests comparing performance across all target odors during training are shown in Fig. 4. Results indicate that TNT and AN were the two targets most difficult to learn, with TNT being significantly more difficult than AN. The remaining targets (including the original Mixture target) overlapped confidence limits with at least one other target. The Mixture target, C4 and C2 were the easiest for the dogs to learn.

3.4. Effect of training paradigm on generalization during acquisition training

We extracted the probability of alerts across all non-reinforced generalization trials for the four odor variants used in training (PE-4 (untagged), PETN-based sheet explosive, CAN26, and French Flake TNT) to determine if one training paradigm led to greater overall responding during generalization tests during acquisition training (see Fig. 5). A logistic mixed effect model indicated there was no interaction between the explosive class and training paradigm (χ^2=2.7, df=2, $p < 0.85$). We therefore removed this term to reduce model complexity. The reduced model indicated a significant effect of training paradigm (χ^2=52.6, df=2, $p < 0.001$). Post-hoc tests show higher generalization rates for Inter-mixed compared to Sequential ($z = 7.14$, $p < 0.001$) and Mixture training ($z = 5.2$, $p < 0.001$). Generalization rates were approaching significance between Mixture training compared to Sequential, with Mixture training having higher generalization rates ($z = 2.2$, $p = 0.068$).

3.5. Effect of training paradigm on post-acquisition generalization

Following acquisition of the eight training odors, dogs completed a series of 14 generalization tests to novel variants of the four explosive

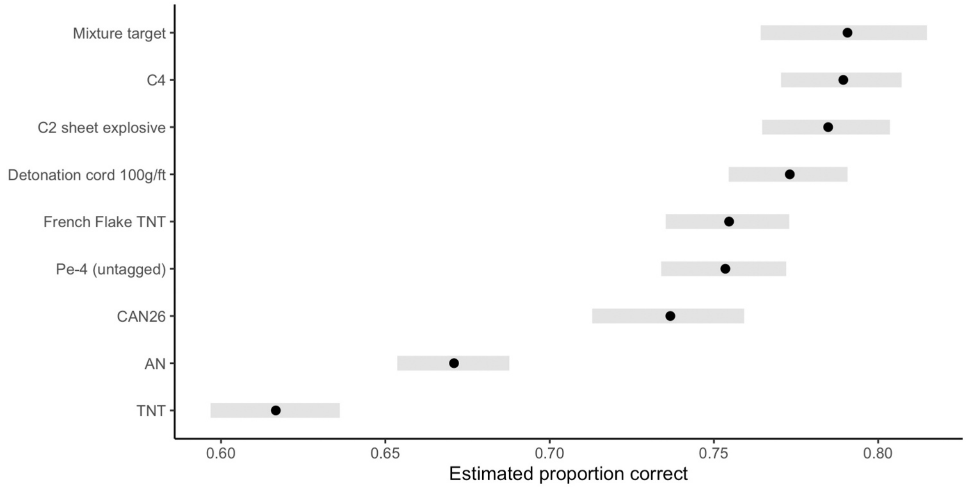

Fig. 4. Estimated proportion correct to a given target during explicit training in steps 3 through 6. The model predicted proportion correct using data across the full acquisition phase, reflecting training difficulty (not generalization). Shows the estimated marginal means and 95 % confidence intervals from the reported mixed effect model.

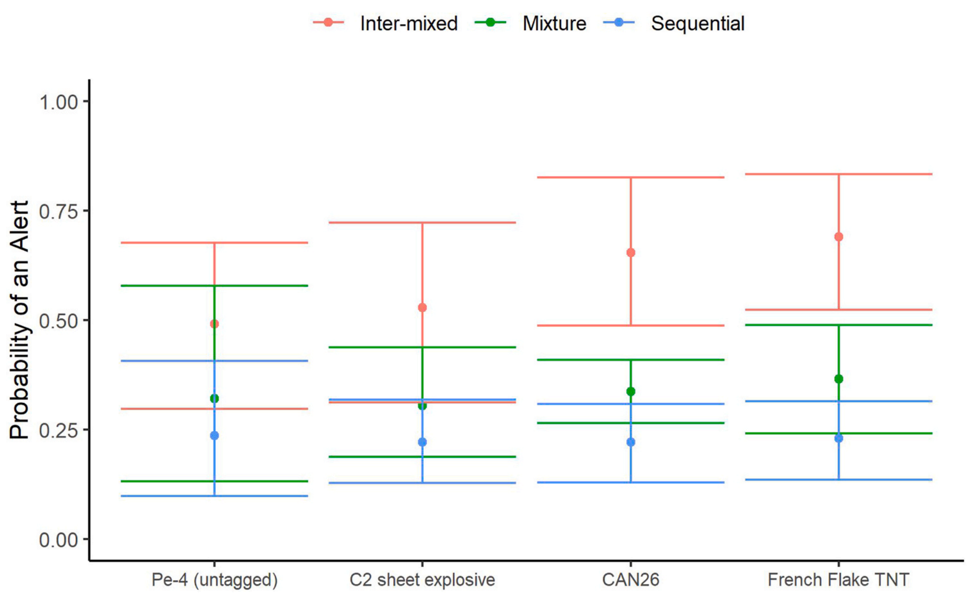

Fig. 5. Relationship between probability of an alert and explosive odor variant identify across all generalization tests. All dogs were naïve to these odor variants upon presentation, although based on training paradigm, not all dogs had equal experience with the original target related to each of these variants.

classes. Fig. 6 shows the probability of alert to the explicit trained target (left of dashed line) and to each generalization variant. We conducted a generalized linear mixed effects model in which proportion of alerts was predicted by explosives class (TNT-based, PETN-based, Plastics-based, AN-based), training paradigm, and their interaction. Random intercepts were included for dogs and institutions.

There was a significant interaction between training paradigm and explosive class (X^2=12.5, DF=6, p = 0.05). Post hoc tests showed that for TNT-based energetics, Inter-mixed training showed greater generalization (72 % alerts) than Sequential (28 % alerts, OR=6.57, z = 2.99, p = 0.004) and Mixture training (25 % alerts, OR=7.34, z = 3.22, p = 0.004). Mixture and Sequential training for TNT did not differ (OR=090, z = -0.17, p = 0.86).

For plastic energetics, Inter-mixed training led to greater generalization (19 % alerts) than Sequential training (3 % alerts, OR=6.93, z = 3.03, p = 0.007), and a trend-level significance from Mixture training (6 % alerts, OR=3.5, z = 2.1, p = 0.06). Mixture and Sequential training for C4 did not differ (OR=2.0, z = 1.0, p = 0.32).

For PETN-based explosives, Inter-mixed training led to more generalization (23 % alerts) than Mixture training (5 % alerts, OR=5.7, z = 2.5, p = 0.036), but not significantly greater than Sequential training (9 % alerts, OR=3.11, z = 1.8, p = 0.12). Mixture and Sequential training did not differ (OR=055, z = -0.80, p = 0.43). For AN variants, Inter-mixed training led to similar generalization (63 % alerts) compared with Sequential (33 % alerts, OR=3.32, z = 1.9, p = 0.17) and Mixture training (54 % alerts, OR=1.34, z-0.47, p = 0.25), and Sequential and Mixture training also led to similar generalization rates (OR=2.5, z = 1.4, p = 0.25).

We observed no responses to untagged C4 (M112), untagged C4 (Bofors), tagged C4 (Bofors) and untagged Flex-X. Other variants resulted in few responses, or poor generalization, see Table 3 for more details.

We conducted an identical analysis using cumulative sniff time as the indicator of generalization (see Fig. 7). We fit a generalized linear mixed effects model in which cumulative sniff time was predicted by explosives class (TNT-based, PETN-based, Plastics, AN-based), training paradigm,

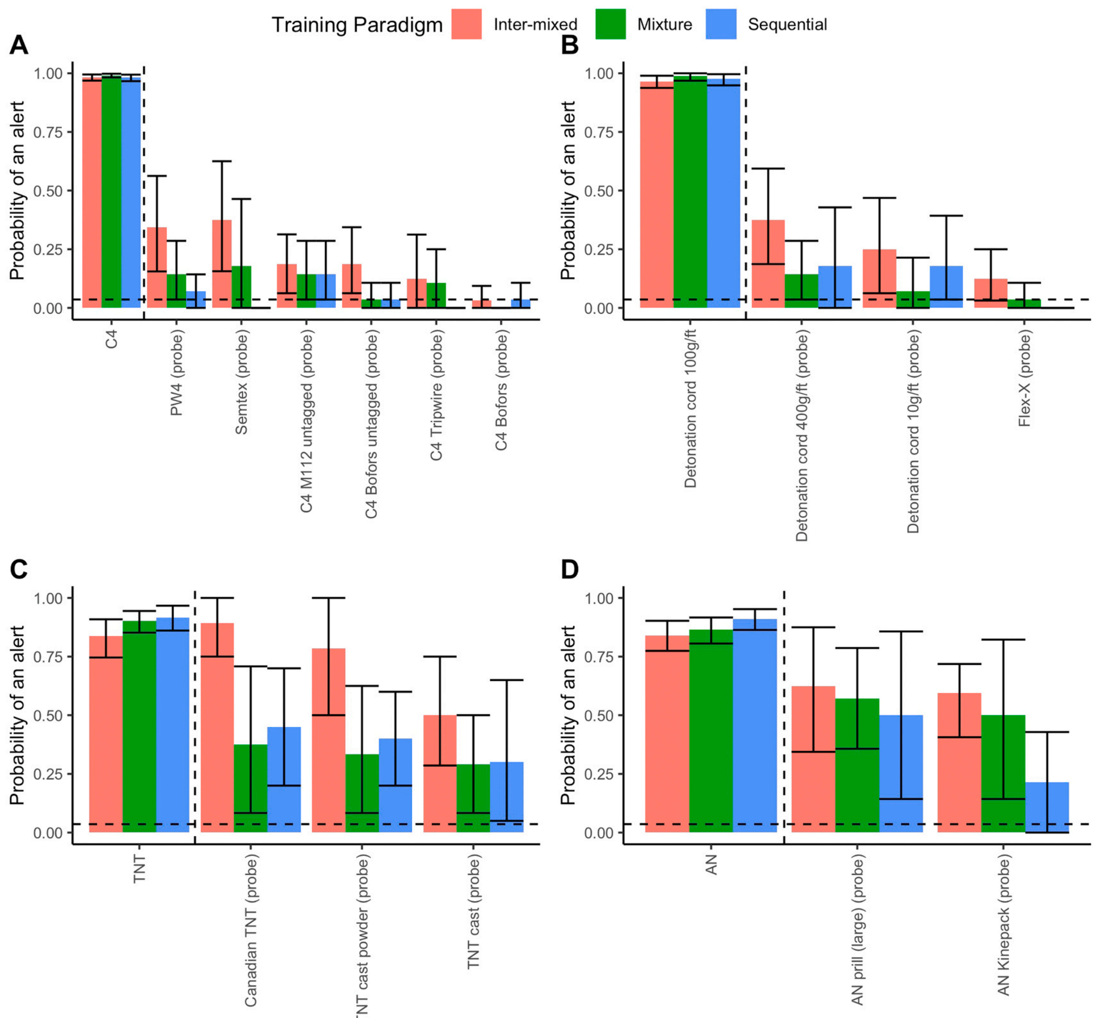

Fig. 6. The effect of training paradigm on probability of a response to novel odors. Each panel shows a different type of explosives, (A) plastics, (B) RDX based, (C) TNT variants (D) AN variants. The horizontal dashed line indicates the average response to distractor odors. The odor to the left of the vertical dashed line is the trained target, while all probe odors are on the right side of each plot. Error bars show the bootstrap estimated 95 % confidence limits.

Table 3
The number of dogs that reached 75 % detection proficiency. Shows the number of dogs (out of number tested) that successfully alerted on at least 3 of the 4 probe trials.

Probe Odor	Inter-mixed	Mixture	Sequential
C4 M112 (untagged)	0 (8)	0 (7)	0 (7)
AN prill (large)	4 (8)	3 (7)	3 (7)
Canadian TNT	6 (7)	2 (6)	2 (5)
TNT cast powder	5 (7)	1 (6)	1 (5)
AN Kinepack	5 (8)	4 (7)	2 (7)
TNT cast	3 (7)	1 (6)	1 (5)
Detonation cord 400 g/ft	1 (8)	0 (7)	1 (7)
PW4	1 (8)	0 (7)	0 (7)
Semtex	2 (8)	1 (7)	0 (7)
Detonation cord 10 g/ft	1 (8)	0 (7)	1 (7)
C4 Bofors (untagged)	0 (8)	0 (7)	0 (7)
C4 Tripwire	1 (8)	0 (7)	0 (7)
Flex-X	0 (8)	0 (7)	0 (7)
C4 Bofors	0 (8)	0 (7)	0 (7)

and their interaction. Random intercepts were included for dog and institution.

There was a significant interaction between training paradigm and explosive class (X^2=20.59, DF=6, p = 0.002). Post hoc tests showed that for TNT-based energetics, Inter-mixed training led to longer sniff times (3.4 s) than Sequential (2.1 s, est=1.3 s, t = 3.87, p < 0.001) and Mixture training (1.8 s, est=1.4 s, t = 4.6, p < 0.001). Mixture and Sequential training for TNT did not differ (est=0.2 s, t = -0.61, p = 0.54).

For plastic energetics, Inter-mixed training led to a trend-level significantly greater sniff times (1.3 s) than Sequential training (0.70 s, est=0.644 s, t = 2.4, p = 0.08), but did not differ from Mixture training (0.93 s, est=0.5 s, t = 1.7, p = 0.15). Mixture and Sequential training did not differ (est=0.17 s, t = 0.63, p = 0.53).

For PETN-based explosives, Inter-mixed training had similar sniff times (1.6 s) compared to Mixture training (1.0 s, est=0.63 s, t = 2.05, p = 0.11), and Sequential training (1.0 s, est=0.55 s, t = 1.8, p = 0.11). Mixture and Sequential training did not differ (est=0.07 s, t = 0.23, p = 0.82).

Finally, for AN variants, Inter-mixed training led to longer sniff times (3.1 s) compared to Mixture (2.0 s, est=0.65 s, t = 1.92, p = 0.09) and significantly longer sniff times compared to Sequential training (2.4 s, est=1.12 s, t = 3.3, p = 0.005). Sequential and Mixture training did not lead to different sniff times (est=0.472 s, t = 1.4, p = 0.17).

Table 3 shows the number of individual dogs for each group that met the 75 % or higher detection threshold in probe trials. This table highlights the number of individual dogs that spontaneously reached detection proficiency for each of the 14 variants. More dogs spontaneously reached detection proficiency for TNT and AN variants, where

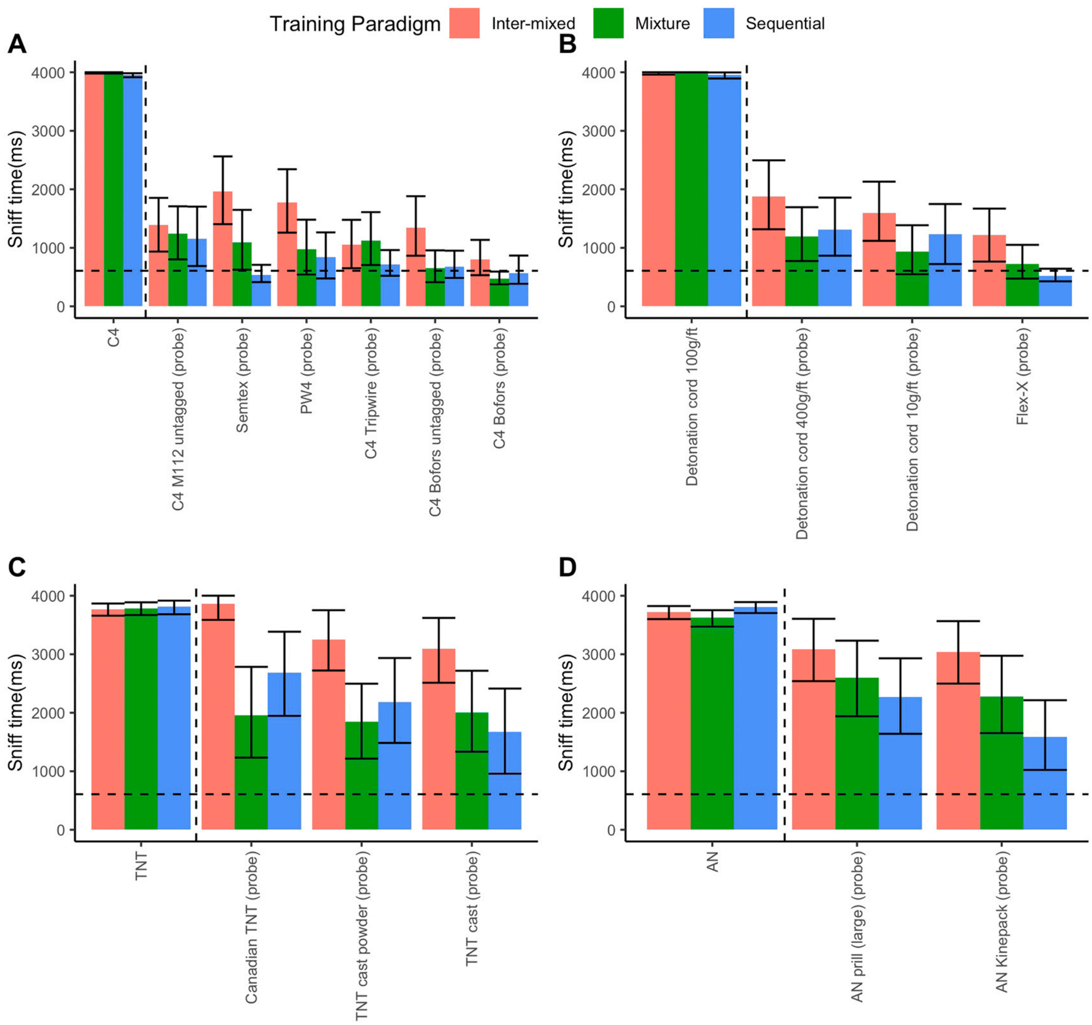

Fig. 7. Effect of training paradigm on cumulative sniff time (ms) on novel target odors. Panels show the average sniff time for each target and probe odor. Plots are split by odor type (A) plastics, (B) PETN-based, (C) TNT variants, (D) AN variants. Dashed horizontal line indicates average sniff time for distractor odors. The odor to the left of the vertical dashed line is the trained target, while all probe odors are on the right side of each panel. Error bars show the bootstrap estimated 95 % confidence limits.

very few showed spontaneous generalization across PETN and many plastic explosive variants.

3.6. Effect of training paradigm on false alarm rate

We also investigated whether the training paradigm influenced the probability of response to novel distractor odors (i.e. false alarms) during generalization tests. Fig. 8 shows overall and odor-specific false alarm rate for each training paradigm. A logistic mixed effects model, which predicted false alerts by training paradigm with a random effect of dog and institution, showed a significant effect of training paradigm (X^2=6.30, DF=2, $p = 0.043$). Post hoc analyses indicate the probability of a false alert for the Inter-mixed training of 7.2 % was significantly higher than the Sequential training false alarm rate of 2.2 % (OR=3.38, $z = 2.47$, $p = 0.04$). The Inter-mixed training false alert rate did not significantly differ compared to the Mixture training false alarm rate of 3.8 % (OR=1.94, $z = 1.41$, $p = 0.33$). Mixture and Sequential training also did not differ in overall false alarm rate (OR=1.74, $z = 1.10$, $p = 0.52$). Inter-mixed dogs showed higher interest in PVC (Fig. 8B), which contains the plasticizer 2-ethyl, 1-hexanol (2E1H) which was present in many of the explosives used for training, indicating perhaps some stimulus control for this VOC.

3.7. VOC-based prediction of canine nose hold time

SPME followed by GC/MS was used to characterize the headspace of each energetic in this study. Using the peak area obtained for each VOC, we conducted Partial Least Square (PLS) regression for each energetic category where observed VOC peak areas predicted cumulative nose hold time.

For plastic explosives cross-validation identified a two-component model that explained a cumulative 38 % of the VOC data and 31 % of the variance in cumulative nose hold time. From this model, we extracted the ten most important VOCs associated with variation in nose hold time, which are shown in order of importance in the panels in Fig. 9. The four most important VOCs were largely identified in variants to which dogs showed minimal response (see Fig. 9). For example, the two most important VOCs were unique to both samples to which little generalization was observed. VOCs found across all plastic explosive samples, such as 2E1H, did not explain the nose hold variance. These common volatiles were found in all plastic explosive samples (those that had high cumulative sniff times, and low), thus they were not predictive of sniff time.

For the TNT, AN and Det Cord energetic categories, the PLS regression explained less than 5 % of the variance in cumulative nose hold time or there were too few variants tested for adequate analysis.

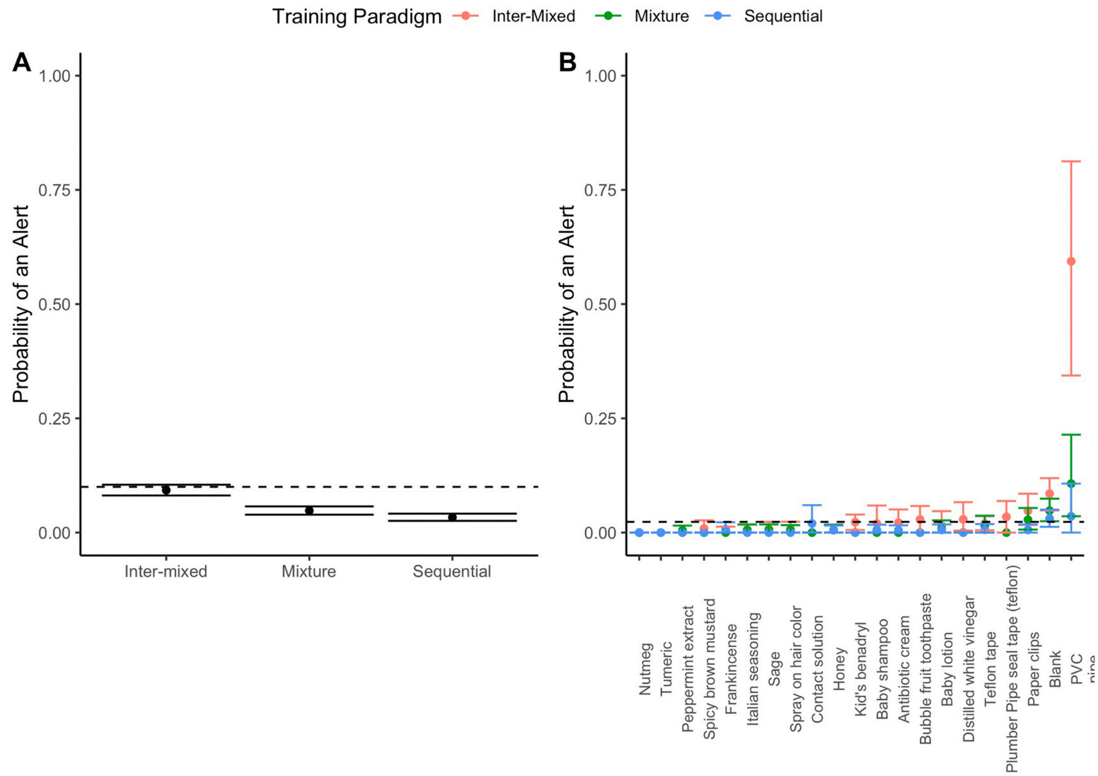

Fig. 8. Effect of training paradigm on probability of an alert to novel distractor odors (i.e., false alarms). A: Shows the probability of a false alert during the large generalization test for the three imprinting paradigms. The horizontal dashed line indicates that 10 % false alert rate. B: Shows the average probability of a response to each of the novel distractors presented during the large generalization test. Dashed line indicates the average probability of a response to a distractor odor (2.34 %). Error bars show the bootstrap estimated 95 % confidence limits.

4. Discussion

This study explored how training paradigm (Mixture, Sequential, Inter-mixed) affects odor detection acquisition time and generalization to novel targets across multiple explosive classes. There was only trend-level significance indicating that the Inter-mixed group learned the eight target odors faster than the Sequential group. There was no difference between the Mixture or Sequential and Mixture and Inter-mixed groups in acquisition rate. Our findings differed from the results of Keep et al. (2021), who found that Mixture training reduced learning time compared to Inter-mixed and Sequential methods. We suspect this difference is related to the specific odors and overshadowing effects in our target mixture.

The Mixture group learned their initial target quickly (an average of 11 sessions to meet criterion), but they largely failed to meet proficiency on most of the individual components of the mixture. Of the dogs in the Mixture group, four dogs generalized to only one target from the mixture (three to C4, one to detonation cord (100 g/ft)), one dog generalized to PETN-sheet explosive, one dog generalized to two targets (C4 and French Flaked TNT), and one failed to generalize to any of the components of the mixture and any variant odors. These low rates of generalization indicate that it would have been just as (if not more) efficient to train them sequentially. These results contradict those shown by Gazit et al. (2021), who found that dogs could discriminate the odors of both three- and five-component mixtures following mixture training. This could be due to differences in the components of the mixture as the odors used by Gazit et al., (2021) were not disclosed due to security reasons, as well as methodological differences in how the mixtures were presented.

We suspect the poor generalization from mixture to components may be partly due to the overshadowing effects of C4. Four dogs generalized to C4 after mixture training, and the remaining three dogs correctly identified C4 in two out of four probe trials (missing the criteria for "spontaneous generalization" by one trial). C4 likely is a more salient stimulus in the component mixture compared to AN and flaked TNT. Dogs have a better detection threshold for C4 compared to AN and TNT (Fernandez et al., 2024), suggesting that relative salience was a contributing factor. We selected targets based on representation of important classes of energetics that dogs are trained to detect. If we had selected only odors of similar intensity, we may have observed more success with the Mixture training. Overshadowing effects have been observed in some previous studies (Kay et al., 2005), but not all (Waggoner et al., 1998).

Training duration also appears to have been driven by odor experience. More naïve dogs at TTU showed that the Inter-mixed group had shorter training durations compared with the Mixture and Sequential groups, but all groups at AU showed similar acquisition durations. It is possible that the more extensive previous odor experience of the dogs at AU led to shorter training durations to novel odors regardless of training paradigm. This suggests that Inter-mixed training could be especially beneficial for less experienced dogs and could expedite initial odor detection training.

Results show that Inter-mixed training led to greater generalization during acquisition and post-acquisition compared to Sequential and Mixture training. The Sequential and Mixture training paradigms did not show significantly different generalization rates during acquisition training. For some odor classes, the generalization improvement could have substantial operational consequences (e.g., 72 % alert rate for TNT variants vs 28 or 25 % alert rate). For other explosive classes (such as Det cord/PETN-based), generalization rate increased from 9 % (Sequential) or 5 % (Mixture) to 23 % (Inter-mixed). This suggests that although Inter-mixed training did improve generalization, the training did not lead to proficiency across this explosive class, where in many cases only one or fewer dogs per group met the 75 % detection criterion for proficiency (Table 3). Explicit training to Det Cord/PETN-based

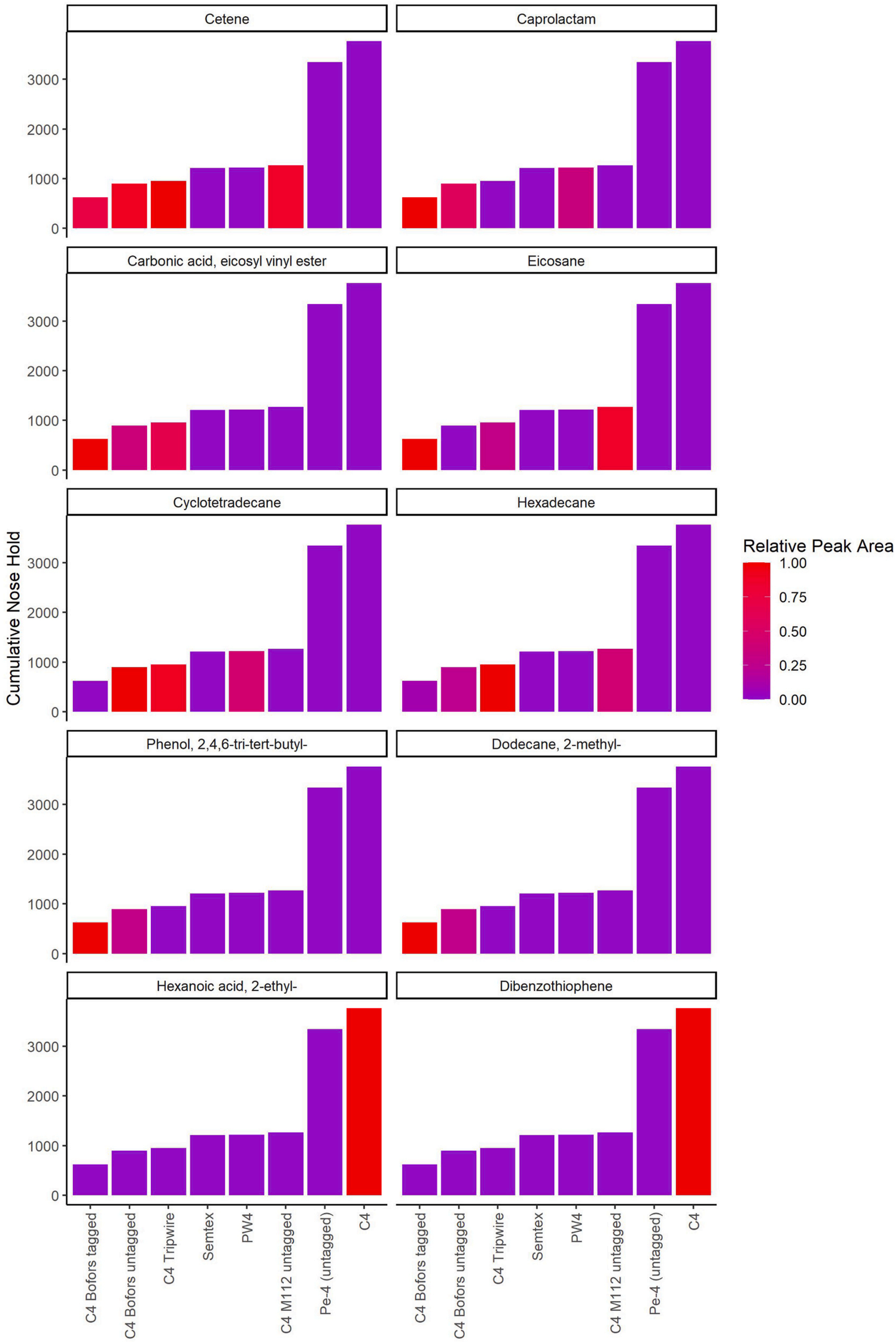

Fig. 9. PLS important VOCs for plastic explosives. Each graph shows a different VOC in order of importance as ranked by the PLS model. Shows the total sniff criterion for each energetic. Plastic explosive variant order on the X axis is based on the average sniff time by dogs. Bar color shows the relative peak abundance that VOC was measured with respect to the energetics evaluated. This was calculated by dividing peak area of a VOC for an explosive divided by the maximum peak area observed for that VOC across the sample set. Thus, a one would indicate that energetic showed the highest concentration for that VOC. A value of zero would indicate that the VOC was not observed for that energetic.

variants is therefore still necessary. Thus, Inter-mixed training with as many exemplars as possible within a class is likely necessary to achieve desired detection proficiency for the variants within this class (Schrier and Brady, 1987).

The mechanism for the improved generalization seen in the Inter-mixed training paradigm is not immediately clear. One potential mechanism for the poorer generalization of the Mixture and Sequential groups, compared to the moderate generalization seen in the Inter-mixed group is an over-training effect in these latter two groups. Previous research indicates that generalization gradients narrow with increased presentations of a reinforced stimulus (S+) (Cleland et al., 2009). Both the Sequential and the Mixture training paradigms (during supplemental training) trained exclusively on one target until they met criterion. This highly focused training on one odor may have produced an over training effect and thus narrowed the generalization gradient around that one stimulus and reduced these groups' ability to generalize to related novel odors. The Inter-mixed group, as they were never singularly trained on one target, avoided over-training, resulting in the observed increased generalization. Another potential explanation for the improved generalization seen in the Inter-mixed group is that dogs in this group were trained with multiple stimuli of varying odor qualities in one session, which may have resulted in a more liberal criterion when presented with a novel stimulus. We know that the Inter-mixed group had a higher false alarm rate, compared to the other two groups, which supports this hypothesis of a more liberal criterion.

We also observed other important generalization failures within the plastic explosive/RDX-based classes. It is important to note that, even after dogs were trained to C4 and Pe4, all training groups showed less than a 20 % response to C4 from a different distributor. This highlights that sourcing from different distributors may produce discrepant responses, from near 100 % alert rate to less than 20 % alert rate, even when using Inter-mixed training. These results support the findings of Dorman et al. (2021), who found that within an explosives class (AN), generalization failures are possible even after training with multiple exemplars.

Inter-mixed training led to a 7.0 % false alarm rate compared to 2.0 % and 3.8 % for Sequential and Mixture training. This suggests that Inter-mixed training changed the tendency of dogs to respond to a novel odor, yielding benefits to generalization to novel targets with draw backs on false alarm rate. The practical importance of this increase in generalization at the risk of an increase in false alarm rate would have to be evaluated by the operational team to assess whether false alarms or misses would be more critical to their operation. For example, increasing generalization to novel TNT variants from 20 % to 72 % could be worth the 2–7 % increase in false alarms. Further, although Inter-mixed training led to increased false alarms, the false alarm rate remained below what has been recommended as an acceptable threshold (American Academy of Forensic Sciences, 2021).

The source of differences in false alarm rates is also noteworthy. Inter-mixed dogs showed higher false alarm rates to blank vials and PVC. PVC uses the plasticizer 2-ethyl-1-hexanol (2E1H) which was observed as a VOC for C4, AN, and French Flake TNT. This may, in part, explain the high response rate to the PVC, or it is also possible PVC is a low salient odor like the blank vials. While 2E1H was not one of the most important volatiles predicting sniff time in our PLS analysis (see Fig. 9), its low importance likely is a result of its prevalence in nearly all samples (both samples that elicited a high sniff time, and those that did not). 2E1H may still be an important volatile in dog recognition of plastic explosives, resulting in the high false rate on PVC, but its commonality among the plastic explosive samples used in the PLS regression, likely deceased its importance in this model. Outside of PVC and blanks, Inter-mixed dogs showed similar false alarm rates to the other training paradigms for all other novel odors (see Fig. 8).

Although dogs are widely used to detect the energetic classes evaluated here, there is little reported knowledge on the difficulty in acquisition between these energetic types. Present acquisition data can be interpreted to suggest that TNT is substantially more difficult to learn than all other energetics trained here, followed by AN. These results agree with previous literature that found dogs have poorer detection sensitivities for TNT and ANs (Fernandez et al., 2024). CAN26, Pe4 and French Flake TNT were similar in acquisition difficulty. C4, 100 g/ft detonation cord and C2-sheet explosive were the easiest odors for dogs to learn. Both CAN26 and French Flake TNT had higher probabilities of a response compared to their corresponding primary target exemplars (AN and American flaked TNT). This may be a result of study design. Once dogs were trained to AN and TNT, it may have been easier for these dogs to learn the variants. Future studies should explore how training to one challenging odor exemplar may influence detection of other challenging odors.

The PLS regression model explained about 30 % of the variance of cumulative nose hold time for plastics. The most important VOCs were odors present in the energetics to which dogs did not generalize. For example, Cetene and Caprolactam were present in variants to which the dogs did not generalize. Cetene is described as "gasoline-like" and caprolactam is described as "disagreeable" (PubChem, 2024), indicating that it may have changed the overall perception of these plastic energetics. Prior research in rodents showed that adding a single odor to a complex odor mixture can greatly impact the similarity of original and adulterated odor mixtures (Barnes et al., 2008). These findings suggest that some generalization failures within plastics may be due to addition of odors rather than the lack of specific volatiles. Common volatiles found in plastic explosives, such as 2E1H (Harper et al., 2005) and cyclohexanone, were not found to predict generalization sniff time. This, however, does not mean that these volatiles are not important for the identification of plastic explosives from distractor odors. The prevalence of 2E1H and cyclohexanone in nearly all samples, both samples that elicited a high sniff time and those that elicited a low sniff time, indicates that they were not useful predictors of *generalization* across plastics in the PLS regression. Had we included VOC analysis with all distractor odors, this model would have likely identified 2E1H and other plasticizers as important markers for plastic explosives. Further, taggant (DMNB) also did not appear as one of the most important drivers of generalization, where dogs generalized almost equally (poorly) to tagged and untagged variations of plastic explosives.

There are important limitations to this work. First, our sample size only covered up to eight dogs per training paradigm. Extending these results to larger and broader populations of dogs would be important, especially given that impacts of Inter-mixed training on generalization were modest. Second, our sample of dogs had varied odor experience, from completely naïve (TTU Cohort 1), with all training occurring within the context of the experiment, to AU Cohort 1, which had extensive prior odor work experience, and TTU Cohort 2, which completed an intra-class explosive study, of this design, prior to completing the present study. Given the small sample, these prior experience effects could not be teased apart here, highlighting the value of future replication.

Another important limitation is the use of non-reinforced probes to assess generalization. Across the many testing sessions, dogs were repeatedly exposed to non-reinforced probes which likely lowered responding. Intermittent schedules of reinforcement were used for the trained targets to help minimize the impact of non-reinforced probes. The use of non-reinforced probes was selected to ensure that spontaneous generalization was being assessed, rather than responses following a reinforced response to an odor. Importantly, dogs did continue to show responses to the various probes across the study period, suggesting that relative comparisons between groups likely remained valid.

When comparing this study to others, such as Gazit et al., (2021), it is important to keep in mind that this project used olfactometers to present odor exemplars to dogs. There may be a difference between an animal sniffing the true material (as done by Keep et al., 2021 and Gazit et al., 2021) and sniffing the headspace of a sample (as done in this study).

Recent work, which employed a scent line up method, found similar results to ours, indicating that method of odor presentation may not affect learning outcomes (Caldicott et al., 2024). Further work is needed in this area, however. Additional work in this area also includes exploring how training explosive odors using an the olfactometer transfers to searching for the true material.

In conclusion, Inter-mixed training led to important improvements in generalization during acquisition and during novel odor tests post-acquisition. Inter-mixed training, however, did not produce adequate detection rates for most novel energetics, although it produced higher rates than the other training paradigms. This suggests that explicit training to as many variants as possible with Inter-mixed training may produce the greatest generalization to untrained variations.

CRediT authorship contribution statement

Jennifer Davis-Miller: Writing – review & editing, Investigation. **Derek Copeland:** Writing – review & editing, Resources, Methodology, Investigation. **Bart Rogers:** Writing – review & editing, Resources, Methodology, Investigation. **Edgar Aviles-Rosa:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Paul Waggoner:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Lucia Lazarowski:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Glenna N. Cupp:** Writing – original draft, Investigation. **Nathaniel Hall:** Writing – review & editing, Visualization, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sarah A. Kane:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Paola A. Prada-Tiedemann:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Conceptualization. Juliet Denapoli: Writing – review & editing, Investigation. **Courtney Collins-Pisano:** Writing – review & editing, Investigation.

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.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.applanim.2025.106638.

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. Animal Cognition 25 (6), 1609–1620.

Aviles-Rosa, E.O., Gallegos, S., Tiedeman, P., Hall, N.J., 2021. An automated canine line-up for detection dog research. Front. Vet. Sci. 1477.

Aviles-Rosa, E.O., Nita, M., Feuerbacher, E., Hall, N.J., 2023. An evaluation of Spotted Lanternfly (Lycorma delicatula) detection dog training and performance. Appl. Anim. Behav. Sci. 258, 105816. https://doi.org/10.1016/j.applanim.2022.105816.

Barnes, D.C., Hofacer, R.D., Zaman, A.R., Rennaker, R.L., Wilson, D.A., 2008. Olfactory perceptual stability and discrimination. Nature neuroscience 11 (12), 1378–1380.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

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) (No Pagination Specified-No Pagination Specified). J. Comp. Psychol.. https://doi.org/10.1037/com0000390.

Cleland, T.A., Narla, V.A., Boudadi, K., 2009. Multiple learning parameters differentially regulate olfactory generalization. Behav. Neurosci. 123 (1), 26–35. https://doi.org/10.1037/a0013991.

DeGreeff, L.E., Peranich, K., 2021. Headspace analysis of ammonium nitrate variants and the effects of differing vapor profiles on canine detection. Forensic Chem. 25, 100342. https://doi.org/10.1016/j.forc.2021.100342.

Dorman, D.C., Foster, M.L., Lazarowski, L., 2021. Training with multiple structurally related odorants fails to improve generalization of ammonium nitrate detection in domesticated dogs (canis familiaris). Article 1. Animals 11 (1). https://doi.org/10.3390/ani11010213.

Fernandez, L.S., Kane, S.A., DeChant, M.T., Prada-Tiedemann, P.A., Hall, N.J., 2024. Environmental effects on explosive detection threshold of domestic dogs. PLOS ONE 19 (9), e0306817. https://doi.org/10.1371/journal.pone.0306817.

Fox, J., Weisberg, S., 2019. _An R Companion to Applied Regression_, Third edition. Sage, Thousand Oaks CA.

Furton, K.G., Myers, L.J., 2001. The scientific foundation and efficacy of the use of canines as chemical detectors for explosives1Invited paper for the special issue of. Talanta 'Methods Explos. Anal. Detect. '. 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.

Gokool, V.A., Crespo-Cajigas, J., Mallikarjun, A., Collins, A., Kane, S.A., Plymouth, V., Nguyen, E., Abella, B.S., Holness, H.K., Furton, K.G., Johnson, A.T.C., Otto, C.M., 2022. The use of biological sensors and instrumental analysis to discriminate COVID-19 odor signatures. Article 11. Biosensors 12 (11). https://doi.org/10.3390/bios12111003.

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.

Harper, R.J., Almirall, J.R., Furton, K.G., 2005. Identification of dominant odor chemicals emanating from explosives for use in developing optimal training aid combinations and mimics for canine detection. Talanta 67 (2), 313–327. https://doi.org/10.1016/j.talanta.2005.05.019.

Kane, S.A., Cupp, G.N., Rangel, M., Medrano, A., Davis-Miller, J., Collins-Pisano, C., Rogers, B., Copeland, D., Lazarowski, L., Waggoner, P., Aviles-Rosa, E.O., Prada-Tiedemann, P.A., Hall, N.J., 2025. The effect of training paradigm on dogs' (Canis familiaris) acquisition and generalization of smokeless powders. Appl. Anim. Behav. Sci. 284, 106527. https://doi.org/10.1016/j.applanim.2025.106527.

Kay, L.M., Crk, T., Thorngate, J., 2005. A redefinition of odor mixture quality. Behav. Neurosci. 119 (3), 726–733. https://doi.org/10.1037/0735-7044.119.3.726.

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. Journal of Statistical Software_ 82 (13), 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.

Lazarowski, L., Foster, M.L., Gruen, M.E., Sherman, B.L., Fish, R.E., Milgram, N.W., Dorman, D.C., 2015. Olfactory discrimination and generalization of ammonium nitrate and structurally related odorants in Labrador retrievers. Anim. Cogn. 18 (6), 1255–1265. https://doi.org/10.1007/s10071-015-0894-9.

Lenth R. (2024). emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package version 1.10.6, .

Moser, A.Y., Bizo, L., Brown, W.Y., 2019. Olfactory generalization in detector dogs. Animals 9 (9), 702.

PubChem. (n.d.). *Caprolactam*. Retrieved March 18, 2024, from ⟨https://pubchem.ncbi.nlm.nih.gov/compound/7768⟩.

Schrier, A.M., Brady, P.M., 1987. Categorization of natural stimuli by monkeys (Macaca mulatta): effects of stimulus set size and modification of exemplars. J. Exp. Psychol.: Anim. Behav. Process. 13, 136–143. https://doi.org/10.1037/0097-7403.13.2.136.

Shellman Francis, V., Holness, H.K., Furton, K.G., 2019. The ability of narcotic detection canines to detect illegal synthetic cathinones (Bath Salts). Front. Vet. Sci. 6. https://doi.org/10.3389/fvets.2019.00098.

Thomas-Danguin, T., Sinding, C., Romagny, S., El Mountassir, F., Atanasova, B., Le Berre, E., Coureaud, G., 2014. The perception of odor objects in everyday life: a review on the processing of odor mixtures. Frontiers in psychology 5, 504.

Waggoner, L.P., Jones, M.H., Williams, M., Johnston, J.M., Edge, C.C., Petrousky, J.A., 1998. Effects of extraneous odors on canine detection. Enforc. Secur. Technol. 3575, 355–362. https://doi.org/10.1117/12.335008.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

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., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. Journal of Open Source Software 4 (43), 1686. https://doi.org/10.21105/joss.01686.