Canine Oil Detection: Generalization and Discrimination of Odor Concentration Report

July 2020

Malloy DeChant and Nathanial Hall

Canine Olfaction Laboratory, Texas Technical University, Lubbock, TX

July 18, 2020

Executive Summary

Canines have demonstrated the ability to detect buried oil but uncertainties remain as to whether a canine can be trained to detect only actionable oil above a certain concentration or quantity and the range of concentrations a dog can detect below the trained concentration. The aim of this project was therefore to (1) assess dogs' spontaneous generalization to lower concentrations of a trained odorant (i.e., the range of lower concentrations dogs alert to without explicit/additional training) and (2) evaluate whether training can suppress responding to a concentration to which dogs previously alerted, while still responding to higher concentrations.

Four detection dogs were trained to alert to a target odorant (0.01 air dilution of the target odorant (~250 ppb vapor concentration)) and spontaneous generalization to a range of concentrations lower than the target was assessed. Dogs generalized (i.e., alerted/responded) to odors within a 10-fold range of the training odorant (≥0.0025; ~62ppb air concentration) but not a 10-fold or greater dilution. Using this generalization gradient, we selected a concentration that dogs showed statistically significant generalization towards and assessed if dogs can learn to successfully discriminate (i.e., stop responding) between the lower concentration odorant (0.0025; ~62ppb air concentration) and the target odorant (0.01, ~250ppb concentration). Dogs successfully discriminated between 0.0025 and 0.01, exceeding 90% accuracy. Interestingly, however, when the second generalization test was conducted, responding at the 0.0025 concentration immediately rebounded. These data suggest that dogs can be trained to actively not respond to concentrations of a trained odorant below a desired threshold. This could be useful for oil-leak detection dogs, in which low concentrations of oil may not present a remediation issue. However, the results suggest that continuous training to not alert to low concentrations may be necessary. Another implication of these results suggests that dogs do not spontaneously generalize to a wide range of concentrations lower than the concentration of the odorant trained. This suggests that if detection of low concentrations is necessary, explicit training at those concentrations will be required.

Table of Contents

Executive Summary 2
Project Overview 4
Introduction 4
Acknowledgements 4
References 12
Figures 12

Project Overview

The primary objective of the 2020 API Canine Oil Detection Study was to address three key uncertainties that previously had not been investigated and specifically to address the following questions:

- Are dogs able to detect weathered and "heavy" oil (that has low to very low volatile organic content)?
- Are dogs able to detect oils to greater depths than previously tested (3 ft - 90 cm)?
- Can a dog be trained to indicate only oil concentrations greater than a certain threshold (for example, 500 ppm in the air above the sample)?

The first two questions were addressed during a set of trials conducted by Owens Coastal Consultants and Chiron K9 on 10-12 February 2020 at the facilities of the Global Training Academy (GTA) near San Antonio, TX. The results of that study have been reported separately. The third question was addressed by the Olfactory Research and Education Laboratory of the Department of Animal and Food Science Canine at Texas Technical University (TTU), Lubbock, TX, during the first semester of 2020 and the results of this component of the study are presented in this report.

Introduction

The highly developed olfactory system of dogs has been utilized in the applications of explosives detection (Furton, 2001; Gazit & Terkel, 2003; Lazarowski et al., 2015; Oxley & Waggoner, 2009), narcotics detection (Bird, 1997; Jezierski et al., 2014), cancer screening and detection (Gordon et al., 2008; McCulloch et al., 2006; E. Moser & McCulloch, 2010; Willis et al., 2004), and conservation related detection (Cablk & Heaton, 2006; Greatbatch et al., 2015; Reed et al., 2011). Therefore, basic research understanding canine olfactory perception could have implications for a wide number of agencies and industries.

Olfactory generalization, or making the same behavioral response to variations of a learned odor (Moser et al., 2019), is a key behavioral phenomenon in need of further research for a variety of detection dog applications. For example, dogs trained to one type of oxidizer common in improvised explosive devices (e.g. ammonium nitrate), also need to respond to chemically related oxidizers (e.g. calcium ammonium nitrate) or when the oxidizer is mixed with other materials; however, dogs often fail to spontaneously generalize across these odor variations (e.g. Hall & Wynne, 2018; Lazarowski et al., 2015; Lazarowski & Dorman, 2014).

Olfactory generalization is an active area of basic research. Animals are required to maintain olfactory perceptual constancy across some variations of complex odors (e.g. no two roses smell the same, but still smell like "rose"), but also be able to discriminate across other variations ("rose" vs. "lilac"). For example, rodents readily generalize (or fail to discriminate) a mixture of 10 components from an identical mixture where one component is removed, but readily discriminate (fail to generalize) when one of the 10 components is adulterated by replacing it with a different chemical (Wilson, 2009).

Predicting the degree or magnitude of generalization to different olfactory stimuli can be challenging, because predicting odor "similarity" from chemical structure is not always

successful. In non-olfactory domains, generalization across a single stimulus dimension (e.g. hertz of tone or wavelength of light) shows a predictable generalization curve (Ghirlanda & Enquist, 2003; Guttman & Kalish, 1956) in which the highest levels of responding occur to the trained stimulus with systematically less responding to increasingly dissimilar stimuli (of higher or lower values) following a gaussian curve. Such generalization curves have rarely been described or tested for olfactory stimuli, except for a few notable exceptions. Generalization across odorants with the same functional group but vary in the length of the carbon chain follow a predictable generalization curve (Hall et al., 2016; Laska & Freyer, 1997; Laska & Hübener, 2001; Laska & Teubner, 1999; Yoder et al., 2014). In addition, honey bees show a predictable pattern of generalization from a trained binary odor mixture to systematically varying ratios of the binary odor mixture (Wright & Smith, 2004).

Interestingly, little work has investigated odor concentration as a dimension for generalization. Perceptual constancy across odor concentration is important for animals to track and locate natural odorants at varying distances from an odor source (Cleland et al., 2012); although, little research has evaluated animals' spontaneous recognition of odors across varying concentrations. In non-chemosensory stimuli, research has previously shown that changes in stimulus intensity produces asymmetric generalization curves, showing typical gaussian generalization for less intense stimuli, but flat and high response for more intense stimuli (Ghirlanda & Enquist, 2003).

Little work, however, has utilized olfactory stimuli. In humans, changes in odor concentration of 100-fold or more can be perceived as producing a different odor quality that is as different as a new chemical (Gross-Isseroff & Lancet, 1988). Further, research suggests that rodents generalize from a trained odor to concentrations 10-fold lower (or greater) than the trained odorant, but show significantly lower responding beyond a 10-fold change in concentration (Cleland et al., 2009, 2012).

To what degree animals can be trained to generalize or discriminate odor concentrations that are within a perceptually similar range has yet to be tested. Further, this information would have important implications for the use and capabilities of detection dogs. For example, in the detection dog handler community it is frequently, but anecdotally, reported that dogs may show great proficiency in detecting a target odor at a concentration they are familiar with, but fail to show any interest when the odor is presented at a different concentration.

Another important application is for when dogs are trained in certain detection applications, such as oil spill remediation. Dogs trained to detect oil from leaking pipes or spills are deployed to rapidly identify and screen for potential oil sources that require remediation. However, it may not be uncommon for trace concentrations of oil to appear when remediation may not be feasible, and thus, dogs need to be trained to ignore such trace, yet detectable, concentrations and respond only to concentrations exceeding some threshold. However, little research has investigated such procedures for training this type of olfactory discrimination although such learning induced changes to generalization gradients have long been known for visual stimuli (Guttman, 1959; Guttman & Kalish, 1956).

The aim of this project was therefore to (1) assess dogs' spontaneous generalization to lower concentrations of a trained odorant (i.e. the range of lower concentrations dogs alert to without explicit/additional training to different concentrations) and (2) evaluate whether training

can suppress responding to a concentration to which dogs spontaneously generalized, while maintaining responding to higher concentrations.

Materials and Methods

Animals. Four mixed breed dogs were utilized and housed at Texas Tech University Canine Olfaction Lab and were participating in a training program to increase adoptability. Dogs' backgrounds were unknown, but all dogs were presumable naïve to detection training. Dogs received twice daily walks, social enrichment, and training. Procedures were reviewed and approved by the Texas Tech University Institutional Animal Care and Use Committee (ACUC# 20010-01).

Air Dilution Olfactometer. To measure thresholds, an air dilution olfactometer that produces one diluted or clean airline was used (see Figure 1). This device takes clean compressed "zero" air and utilizes Alicat® gas mass air flow controllers to do a serial air dilution of an odorant held in a glass vial designed for volatile organic compound sampling that rested in a temperature-controlled bath at 38°C. All odor whetted parts are made of Polytetrafluoroethylene (PTFE, Teflon), Stainless Steel 316 and glass per olfactometry standards (EN 13725). A final stimulus flow rate of 10 L/min was used for all odor concentrations and clean air stimuli.

Odorant. All dogs were trained to detect isoamyl acetate (CAS #123-92-2). To reduce contamination within the olfactometer, the odorant was pre-diluted in mineral oil to a liquid dilution of $10^{-2.5}$ v/v in mineral oil.

Odorant delivery validation. To validate odorant delivery prior to the experiment, a 10^{-2} v/v dilution of isoamyl acetate was used to generate odorant stimuli to be read by a photo-ionization detector (PID, 200b miniPID, Aurora Scientific®). The miniPID was placed in the odor port (medium gain, high pump) where a dog would place its nose. The mean of 200 samples at each of five different concentrations (0.1, 0.07, 0.05, 0.025, 0.01) was taken during the stable phase of odor activation to create a calibration curve using the mean PID voltage reading with the theoretical vapor dilution generated. The range of vapor dilutions used was higher than the range of concentrations presented to the dog, so that the odorant was sufficiently concentrated to be readable by the PID.

Preliminary Training. Dogs were trained on a Go/No-Go task. Dogs were presented with a PTFE odor port approximately 51cm from the ground, an exhaust line to remove the odorant, infrared sensors to detect dog presence, and a retractable lever that was present during the entire session. An automated cover restricted access to the port between trials, such that the dogs were only able to access the odor port during designated periods. The intertrial interval (ITI) was 20s and the automated cover prevented access to the odor port during the ITI. An exhaust fan evacuated residual odor during the ITI and the olfactometer initiated the odor (or no odor) presentation for the subsequent trial during the ITI, such that at the end of the ITI, the stable odor concentration was ready for the next trial.

An automated feeder rewarded dogs for correct responses. The target odor was presented as a 0.01 air dilution of the diluted isoamyl acetate odorant ($10^{-2.5}$ v/v). Dogs 'alerted' by holding their nose in the odor port for 3s, measured by infrared beams, or by pressing the lever.

Three dogs tended to make a "nose hold" response whereas as one dog made mostly lever presses. Both responses were counted as "indication" responses. The "no-go" response was comprised by not holding their nose in the odor port for more than 3 continuous seconds and by not pressing the lever.

Dogs worked with the same handlers and trained and tested in the same experimental room. Dogs were initially trained to the automated go/no-go device by placing a piece of food in the odor port until the dog learned to insert their nose. Afterwards, shaping to the target odor, isoamyl acetate, with a clicker and food reward was repeated until the dog indicated on the odor in the port. Dogs were taught to 'alert' by either holding their nose in the odor port for 3 sec or by pressing the lever. The final response was determined simply by which the dog learned first. Dogs were trained to press the lever by placing a piece of food on the lever until the dog learned to press with its nose.

Training. Once dogs learned to operate the olfactometer independently (without experimenter assistance and experimenter outside of the testing area), Dogs were trained in 40 trial sessions to discriminate clean air from a 0.01 air dilution of the odorant, completely under computer control. Correct responses (hits and correct rejections) were reinforced with delivery of a treat from an automated feeder. Misses and false alarms were not reinforced. If dogs made two consecutive incorrect responses of the same type (miss or false alarm) without an intervening correct response, a correction was conducted. Instead of terminating the trial without food for the incorrect response, the olfactometer continued to present the stimulus until the dog made the correct response (made an alert or rejection response). The trial was still scored as incorrect, but these trials were used to maintain motivation if multiple incorrect responses were made. Once dogs achieved 75% accuracy or higher for two consecutive sessions (binomial test, $p<0.0001$), dogs moved onto intermittent reinforcement training.

Intermittent Reinforcement Training. After successfully reaching criterion with continuous reinforcement, dogs were adjusted to an intermittent schedule of reinforcement in which the probability of reinforcement for any correct response was uniform at 0.8. Dogs received two consecutive sessions of intermittent reinforcement training to adjust dogs to non-continuous reinforcement for the generalization tests where non-reinforced probe trials will be presented.

Generalization Test 1. Immediately following intermittent reinforcement training, dogs underwent eight 40-trial sessions of generalization testing. Generalization testing was identical to training except that 20% of the trials (8 of the 40 trials) were non-reinforced probe trials to measure spontaneous generalization to different concentrations. No corrections were made for responses on probe trials, and simply terminated the trial. On half of the sessions, dogs received two trials of each of the following probe air dilutions of the target odor (isoamyl acetate diluted to $10^{-2.5}$ v/v in mineral oil): 0.003, 0.001, 0.0003, 0.0001. Assuming a vapor pressure of ~6 mmHg (PubChem, Isoamyl acetate, CID=31276), this would equate to a vapor concentration of ~250 ppb, 75ppb, 24ppb, 7 ppb, and 2.5 ppb, respectively. On the remaining half of sessions, two trials of the following air dilutions were tested: 0.0075, 0.005, 0.0025, 0.00075 [~187 ppb, 124 ppb, 62 ppb, 18 ppb respectively]. Testing was alternated between the two sets of probe dilutions for all 8 sessions.

Concentration Discrimination. Following Generalization test 1, we identified the lowest concentration dogs were significantly more likely to alert toward than the clean air stimulus. To evaluate if dogs can learn to accurately discriminate between an odor to which they spontaneously generalize to (0.0025 dilution: see Results) from the trained concentration (0.01), six training sessions were conducted in which dogs were required to discriminate these two concentrations. The 0.0025 dilution replaced all clean air trials (0.01 vs. 0.0025) and only rejection responses toward the 0.0025 concentration were reinforced and alert responses to 0.01.

Generalization Test 2. Following the six concentration training sessions eight generalization test sessions were conducted identically to Generalization Test 1.

Compound Generalization Test and Discrimination. Following Generalization Test 2, to evaluate if alerting to low concentration odors can be further suppressed, another series of eight generalization test sessions were conducted. However, different from the previous generalization tests, the contingencies in effect for Concentration Discrimination training (0.01 vs. 0.0025) remained in effect for 80% of the trials during the sessions. Alerts to a concentration of 0.0025 were therefore not reinforced and no-go responses to 0.0025 were. Generalization probes to 0.0025 were not conducted (because they were explicitly reinforced), and instead replaced with the concentration of 0 (clean air) to assess responding to clean air.

Statistical Analysis

A generalized linear mixed-effect model for binomial data (response or no response) using a logit-link was used throughout. For all models, Dog ID was used as a random effect. The lme4 package (Bates et al., 2015) in R (R version 3.5.1, www. r.project.org; R Core Team, 2018) was used to fit models. Post-hoc tests were conducted using the lsmeans package (Lenth, 2016).

Results

The air dilution olfactometer calibration curve is shown in Figure 2. The PID output was highly related to systematic changes in the air dilution olfactometer within the dilution range used ($t=804.8$, $p<0.0001$; $R^2=0.998$). However, the PID showed little detection of the 0.01 concentration, indicating the concentrations used by the dogs for this study were below the detection limits of the PID, and therefore, accuracy in odor delivery for the concentrations presented to the dogs must be assumed from extrapolation from the data within the sensor's limits of detection.

Figure 3 shows dogs' spontaneous generalization to a 100-fold range of concentrations below the trained target concentration. A generalized-linear mixed-effect model was used compare the probability of an alert at each probed dilution in comparison to the clean air (negative) stimulus. Overall, there was a significant effect to odor concentration on dogs' probability of making an alert response ($\chi^2=750$, $df=9$, $p<0.001$). Dunnet-adjusted post-hoc tests, comparing the probability of response at each concentration to clean air, indicates that dogs were significantly more likely to alert to the trained odorant at a 0.01 dilution ($z=19.58$, $p<0.001$), the probe dilution 0.0075 ($z=8.07$, $p<0.001$), probe dilution 0.005 ($z=6.97$, $p<0.001$), probe dilution 0.003 ($z=4.01$, $p<0.001$) and probe dilution 0.0025 ($z=3.13$, $p=0.01$). Dogs, however, alerted to the remainder lower concentrations at the same rate as clean air (0.001, 0.00075, $10^{-3.5}$, 10^{-4}, all $p>0.05$). Thus, dogs spontaneously alerted to any concentration within a

10-fold dilution of the target odor but failed to respond to the 10-fold dilution itself and any dilution lower.

Given that a concentration of 0.0025 was the lowest concentration dogs spontaneously alerted towards, greater than clean air, Concentration Discrimination training was initiated in which alerts to 0.01 dilution were still reinforced but alerts to 0.0025 were not (but no-go responses to 0.0025 were reinforced). Figure 4 shows the mean probability of alerting to the 0.0025 concentration and mean overall accuracy for the contingencies in effect across the six sessions of Concentration Discrimination training. A generalized linear mixed effect model predicting alert probability to the 0.0025 concentration across the six sessions, indicates dogs showed lowering probability in alerting to the 0.0025 concentration across sessions (effect of day: $z=3.00$, $p<0.001$). In addition, dogs showed improvement in overall accuracy with the contingencies across training days reaching over 90% accuracy alerting to the 0.01 concentration and not alerting to 0.0025, but the effect of day did not reach statistical significance ($z=1.07$, $p=0.28$).

Following Concentration Discrimination, the generalization test was repeated (Generalization Test 2; see Figure 5). A generalized linear mixed-effect model, predicting response probability as a function of odor concentration, the test number (Generalization Test 1 or Test 2) and their interaction, indicated that there was not a significant interaction ($\chi^2=13.92$, $df=9$, $p=0.13$). However, to break down comparison between the two generalization tests at each concentration, Tukey-adjusted post hoc tests comparing the probability of an alert between Generalization Test 1 and 2 at each concentration were run. Post-hoc tests indicated that in Test 2, dogs were significantly less likely to false alert on the clean air ($z=3.06$, $p=0.002$). The probability to alert to all other concentrations, however, was not significantly different between Test 1 and 2, including the 0.0025 concentration ($z=1.053$, $p=0.29$) in which dogs were successfully taught to not alert toward in the immediately previous Concentration Discrimination sessions.

Following Generalization Test 2, we evaluated whether generalization might be relative to the contingencies in effect during the sessions (i.e. is a clean air stimulus the no-go stimulus or another concentration). Dogs' generalization was next evaluated in identical generalization sessions except that 0.0025 was used as the no-go stimulus, thereby combining contingencies of the Concentration Discrimination and Generalization Test sessions.

Figure 6 shows the mean probability of response to each concentration during the Compound Generalization and Discrimination and during the last two sessions of the Concentration Discrimination training. In both these conditions, the same contingency was in effect (alerts to 0.01 reinforced and alerts to 0.0025 not reinforced). A generalized-linear mixed effect model, predicting response probability as a function of odor concentration, session type and their interaction, indicated there was no significant interaction between the odor concentration and session ($\chi^2=2.77$, $df=1$, $p=0.10$), or main effect of session ($\chi^2=0.68$, $df=1$, $p=0.40$), indicating responding was similar in both conditions.

Figure 7 shows the results across all generalization sessions (Generalization Test 1, Generalization Test 2, and Compound Generalization and Discrimination). A generalized linear mixed-effect model was fit in which the probability of an alert was predicted by odor concentration, the session type (Generalization Test 1, Test 2, or Compound Generalization and

discrimination), and their interaction. There was a significant interaction between the session type and odor concentration (χ^2=44.45, df=18, p=0.001). Tukey-adjusted pairwise post-hoc tests at each concentration indicate that at the target odor concentration (0.01), the probability of an alert was lower in the Compound Discrimination and Generalization phase than in the Generalization Test 2 phase (z=3.17, p=0.004), and the Generalization Test 1 phase (z=4.38, p<0.001). At the next lower concentration (0.0075), there were no statistically significant differences between the three phases. At a concentration of 0.005, the probability of an alert was lower in the Compound Discrimination and Generalization sessions than the Generalization Test 2 (z=3.03, p<0.001), and the first Generalization Test (z=3.03, p=0.007). At the next lower concentration (0.003), a similar result held, such that the probability of an alert tended to be lower in the Compound Discrimination and Generalization sessions than the Generalization Test 1 (z=1.82, p=0.16), and was almost significantly lower than in Generalization Test 2 (z=2.28, p=0.06). At the concentration used for training dogs not to alert (0.0025), dogs showed significantly less probability to respond in the Compound Discrimination and Generalization session compared to Generalization Test 1 (z=3.62, p<0.001) and Generalization Test 2 (z=7.13, p<0.0001). For all the remaining lower concentration probed concentrations, post-hoc tests revealed no significant differences between the three sessions types, except for the clean air (no odor) concentration. For the clean air stimulus, the lowest probability to alert (i.e. false alert), was in the Compound Discrimination and Generalization condition but this was not significantly different from Generalization Test 1 (z=0.02, p=0.99) and Generalization Test 2 (z=0.02, p=0.99). As previously shown, however, the probability of a response to clean air was lower in Generalization Test 2 than Generalization Test 1 (z=3.06, p=0.006).

Discussion

The results indicated that dogs spontaneously generalize in non-reinforced probes only within a 10-fold range of the trained odor concentration. At a 10-fold dilution and lower, the probability of response was no greater than it was to clean air. This is an important finding as it suggests that dogs only spontaneously generalize to a limited range of an odor concentration without explicit training. One possible explanation is that perhaps dogs could simply not detect a concentration <10-fold dilution of the target odor. However, assuming a vapor pressure of ~6 mmHg (PubChem, Isoamyl acetate, CID=31276), the anticipated vapor concentration (assuming ideal conditions) would be approximately 25 ppb for the 10-fold dilution of the target odorant. Given that recent work has shown amyl acetate thresholds to be close to 1 ppt for dogs (Concha et al., 2019; Walker et al., 2006), our test concentrations should be well above the dog detection limits. This suggests that the low probability of response was a generalization failure, not a sensitivity failure. We selected the concentration range used in this study to optimize dog detection (sufficiently salient) without being too strong as to produce nuisance contamination in the system. Future work, however, can evaluate whether the same results are observed at varying starting concentrations for the trained odorant.

Results from the Concentration Discrimination sessions indicate dogs can readily learn to discriminate a 0.0025 dilution from a 0.01 dilution (a difference of only 0.0075), achieving greater than 90% accuracy within six sessions. Dogs successfully made go responses to the 0.01 dilution, and no-go responses to the 0.0025 dilution. Interestingly, however, despite successfully making nearly all no-go responses to the 0.0025 concentration, when the second generalization test was conducted, responding at the 0.0025 concentration immediately rebounded. This

suggests that the training context (i.e. the dilutions used within a session and their contingencies) may have an important impact on generalization responding.

By inserting probe trials within a session in which 0.01 is discriminated from 0, dogs maybe more likely to respond to any stimulus smelling close to the 0.01 target. In other sensory domains, generalization has been demonstrated to be under relative stimulus control vs. absolute control of the stimulus. For example, the early generalization gradient work demonstrated that when an extinction stimulus in introduced (i.e. a stimulus to which alerts are not reinforced), responding is shifted away from the extinction (non-reinforced) stimulus, leading maximum responding to values exceeding the trained stimulus, termed a "peak shift" (i.e. not the exact value of the reinforced stimulus; Guttman, 1959; Hanson, 1959; Purtle, 1973). It appears a similar process may occur with respect to olfactory stimuli.

In the final session, in which probe trials were inserted within sessions in which a 0.01 dilution was discriminated from the 0.0025 concentration, we found that suppression of responding to the 0.0025 concentration was maintained, and responding to 0.01 was maintained, although slightly lower than observed in the other generalization tests. This could suggest that had we tested generalization to higher concentrations, we may have observed a typical peak shift.

Importantly, however, this does suggest that dogs can be trained to alert to concentrations greater than some threshold, and not respond to lower concentrations that the dog perceives as being similar to the target concentration. These results further suggest that successful application of this may require constant training with the low concentration odorant as a no-go stimulus to maintain suppression of responding and prevent alerting to the low concentration from rebounding as it did in the second generalization test.

Another interesting finding is that false alarms decreased after Concentration Discrimination training. This may result from training dogs not to respond to the low, but detectable concentration odor (0.0025), which made the clean air stimulus even more perceptible as a no-go stimulus.

There are several important limitations to the present findings. The first is the required inherent training order in the sessions. This makes it challenging to determine if the decrease in false alarms may simply be a maturation/experience effect across time or may reflect behavioral changes due to training the 0.0025 concentration as a no-go stimulus. Ideally, we would have further alternated between generalization test sessions and the compound discrimination and generalization sessions to assess steady state responding, reducing experience effects. However, the 2020 SARS-COV2 pandemic prevented further testing between the conditions, but this would remain an important follow-up.

Another limitation is that we explicitly reinforced correct no-go responses. This differs from typical peak-shift protocols that suppress responding with extinction (i.e. simply not reinforcing any response except an alert). We opted not to use this option due to concern that non-reinforced probe trials would thin the reinforcement schedule too much to maintain responding. We do not expect this would have changed the direction of any effects observed but may have more effectively suppressed alerting to the low concentration odorant, given the additional reinforcer to not respond in addition to extinction of any alerts.

The present results also have important implications for detection dogs. It suggests that dogs can be trained to actively not respond to concentrations of a trained odorant below a desired threshold. This could be useful for oil-leak detection dogs, in which detection of low concentrations of oil may not present a remediation issue. However, the results suggest that continuous training to not alert to low concentrations may be necessary. Another implication of these results suggests that dogs do not spontaneously generalize to a wide range of concentrations lower than the concentration of the odorant trained. This suggests that if low concentrations are necessary to detect, explicit training at those concentrations will be required. As noted previously, the range of concentrations selected for this study were done so to optimize dog detection (sufficiently salient) without being too strong as to produce nuisance contamination in the system, but it would be anticipated that similar results would hold for a wide range of concentrations that could be selected as the threshold, to which alerts are only desired for concentrations greater than.

Another important implication for detection dog training is in regard to potential training problems. If odor samples are not prepared carefully, and samples that a trainer may think are "clean" are incidentally contaminated with a significant amount of the odorant, this research suggests dogs can learn to not alert to the contaminated container and still respond to the higher concentration target. From the trainer's perspective, the dog may seem proficient in detection. Critically, however, the dog may explicitly ignore significant concentrations of the odorant that resemble the contamination concentration, which may only be slightly lower than the trained odor concentration (the present study indicates dogs can successfully ignore a 0.25 dilution of the trained odor). This may lead to poorer than expected performance in operational scenarios.

In conclusion, the present results suggest dogs may only spontaneously generalize to concentrations less than a 10-fold dilution of the training odorant. Further, dogs can learn to successfully discriminate between a higher concentration odorant and a 0.25 dilution of that odorant within 6 training sessions. Lastly, dogs can successfully learn to only alert to concentrations greater than a certain threshold, but this may need to be done within a context of consistent training to not respond to the threshold concentration.

Acknowledgements

The study was funded by the American Petroleum Industry (API) through the Oil Spill Science and Technology Working Group. P. Bunker provided technical advice and assistance on the study design and the project was managed by Owens Coastal Consultants, Ltd.

References

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *1*(67), 1–48.

Bird, R. C. (1997). An examination of the training and reliability of the narcotics detection dog. *Ky. Law Jour.*, *85*, 405–433.

Cablk, M. E., & Heaton, J. S. (2006). Accuracy and Reliability of Dogs in Surveying for Desert Tortoise (Gopherus Agassizii). *Ecological Applications*, *16*(5), 1926–1935. https://doi.org/10.1890/1051-0761(2006)016[1926:AARODI]2.0.CO;2

Cleland, T. A., Chen, S.-Y. T., Hozer, K. W., Ukatu, H. N., Wong, K. J., & Zheng, F. (2012). Sequential mechanisms underlying concentration invariance in biological olfaction. *Frontiers in Neuroengineering*, *4*. https://doi.org/10.3389/fneng.2011.00021

Cleland, T. A., Narla, V. A., & Boudadi, K. (2009). Multiple Learning Parameters Differentially Regulate Olfactory Generalization. *Behavioral Neuroscience*, *123*(1), 26–35. https://doi.org/10.1037/a0013991

Concha, A. R., Guest, C. M., Harris, R., Pike, T. W., Feugier, A., Zulch, H., & Mills, D. S. (2019). Canine Olfactory Thresholds to Amyl Acetate in a Biomedical Detection Scenario. *Frontiers in Veterinary Science*, *5*. https://doi.org/10.3389/fvets.2018.00345

Furton, K. (2001). The scientific foundation and efficacy of the use of canines as chemical detectors for explosives. *Talanta*, *54*(3), 487–500. https://doi.org/10.1016/S0039-9140(00)00546-4

Gazit, I., & Terkel, J. (2003). Explosives detection by sniffer dogs following strenuous physical activity. *Applied Animal Behaviour Science*, *81*(2), 149–161. https://doi.org/10.1016/S0168-1591(02)00274-5

Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour*, *66*(1), 15–36. https://doi.org/10.1006/anbe.2003.2174

Gordon, R. T., Schatz, C. B., Myers, L. J., Kosty, M., Gonczy, C., Kroener, J., Tran, M., Kurtzhals, P., Heath, S., Koziol, J. A., Arthur, N., Gabriel, M., Hemping, J., Hemping, G., Nesbitt, S., Tucker-Clark, L., & Zaayer, J. (2008). The Use of Canines in the Detection of Human Cancers. *The Journal of Alternative and Complementary Medicine*, *14*(1), 61–67. https://doi.org/10.1089/acm.2006.6408

Greatbatch, I., Gosling, R. J., & Allen, S. (2015). Quantifying Search Dog Effectiveness in a Terrestrial Search and Rescue Environment. *Wilderness & Environmental Medicine*, *26*(3), 327–334. https://doi.org/10.1016/j.wem.2015.02.009

Gross-Isseroff, R., & Lancet, D. (1988). Concentration-dependent changes of perceived odor quality. *Chemical Senses*, *13*(2), 191–204. https://doi.org/10.1093/chemse/13.2.191

Guttman, N. (1959). Generalization gradients around stimuli associated with different reinforcement schedules. *Journal of Experimental Psychology*, *58*(5), 335–340. https://doi.org/10.1037/h0045679

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, *51*(1), 79–88.

Hall, N. J., Collada, A., Smith, D. W., & Wynne, C. D. L. (2016). Performance of domestic dogs on an olfactory discrimination of a homologous series of alcohols. *Applied Animal Behaviour Science*, *178*, 1–6. https://doi.org/10.1016/j.applanim.2016.03.016

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

Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, *58*(5), 321–334. https://doi.org/10.1037/h0042606

Jezierski, T., Adamkiewicz, E., Walczak, M., Sobczyńska, M., Górecka-Bruzda, A., Ensminger, J., & Papet, E. (2014). Efficacy of drug detection by fully-trained police dogs varies by breed, training level, type of drug and search environment. *Forensic Science International*, *237*, 112–118. https://doi.org/10.1016/j.forsciint.2014.01.013

Laska, M., & Freyer, D. (1997). Olfactory Discrimination Ability for Aliphatic Esters in Squirrel Monkeys and Humans. *Chemical Senses*, *22*(4), 457–465. https://doi.org/10.1093/chemse/22.4.457

Laska, M., & Hübener, F. (2001). Olfactory discrimination ability for homologous series of aliphatic ketones and acetic esters. *Behavioural Brain Research*, *119*(2), 193–201. https://doi.org/10.1016/S0166-4328(00)00348-X

Laska, M., & Teubner, P. (1999). Olfactory Discrimination Ability for Homologous Series of Aliphatic Alcohols and Aldehydes. *Chemical Senses*, *24*(3), 263–270. https://doi.org/10.1093/chemse/24.3.263

Lazarowski, L., & Dorman, D. C. (2014). Explosives detection by military working dogs: Olfactory generalization from components to mixtures. *Applied Animal Behaviour Science*, *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. *Animal Cognition*, *18*(6), 1255–1265. https://doi.org/10.1007/s10071-015-0894-9

Lenth, R. L. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *1*(69), 1–33.

McCulloch, M., Jezierski, T., Broffman, M., Hubbard, A., Turner, K., & Janecki, T. (2006). Diagnostic Accuracy of Canine Scent Detection in Early- and Late-Stage Lung and Breast Cancers. *Integrative Cancer Therapies*, *5*(1), 30–39. https://doi.org/10.1177/1534735405285096

Moser, A. Y., Bizo, L., & Brown, W. Y. (2019). Olfactory Generalization in Detector Dogs. *Animals*, *9*(9), 702. https://doi.org/10.3390/ani9090702

Moser, E., & McCulloch, M. (2010). Canine scent detection of human cancers: A review of methods and accuracy. *Journal of Veterinary Behavior*, *5*(3), 145–152. https://doi.org/10.1016/j.jveb.2010.01.002

Oxley, J. C., & Waggoner, L. P. (2009). Chapter 3- Detection of Explosives by Dogs. In *Aspects of Explosives Detection* (pp. 27–40). Elsevier.

Purtle, R. B. (1973). Peak shift: A review. *Psychological Bulletin*, *80*(5), 408–421. https://doi.org/10.1037/h0035233

Reed, S. E., Bidlack, A. L., Hurt, A., & Getz, W. M. (2011). Detection distance and environmental factors in conservation detection dog surveys. *The Journal of Wildlife Management*, *75*(1), 243–251. https://doi.org/10.1002/jwmg.8

Walker, D. B., Walker, J. C., Cavnar, P. J., Taylor, J. L., Pickel, D. H., Hall, S. B., & Suarez, J. C. (2006). Naturalistic quantification of canine olfactory sensitivity. *Applied Animal Behaviour Science*, *97*(2), 241–254.

Willis, C. M., Church, S. M., Guest, C. M., Cook, W. A., McCarthy, N., Bransbury, A. J., Church, M. R. T., & Church, J. C. T. (2004). *Olfactory detection of human bladder cancer by dogs: Proof of principle study*. *329*, 6.

Wilson, D. A. (2009). Pattern Separation and Completion in Olfaction. In T. E. Finger (Ed.), *International Symposium on Olfaction and Taste* (Vol. 1170, pp. 306–312). Blackwell Publishing.

Wright, G. A., & Smith, B. H. (2004). Variation in complex olfactory stimuli and its influence on odour recognition. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *271*(1535), 147–152. https://doi.org/10.1098/rspb.2003.2590

Yoder, W. M., Setlow, B., Bizon, J. L., & Smith, D. W. (2014). Characterizing Olfactory Perceptual Similarity Using Carbon Chain Discrimination in Fischer 344 Rats. *Chemical Senses*, *39*(4), 323–331. https://doi.org/10.1093/chemse/bju001

Figures

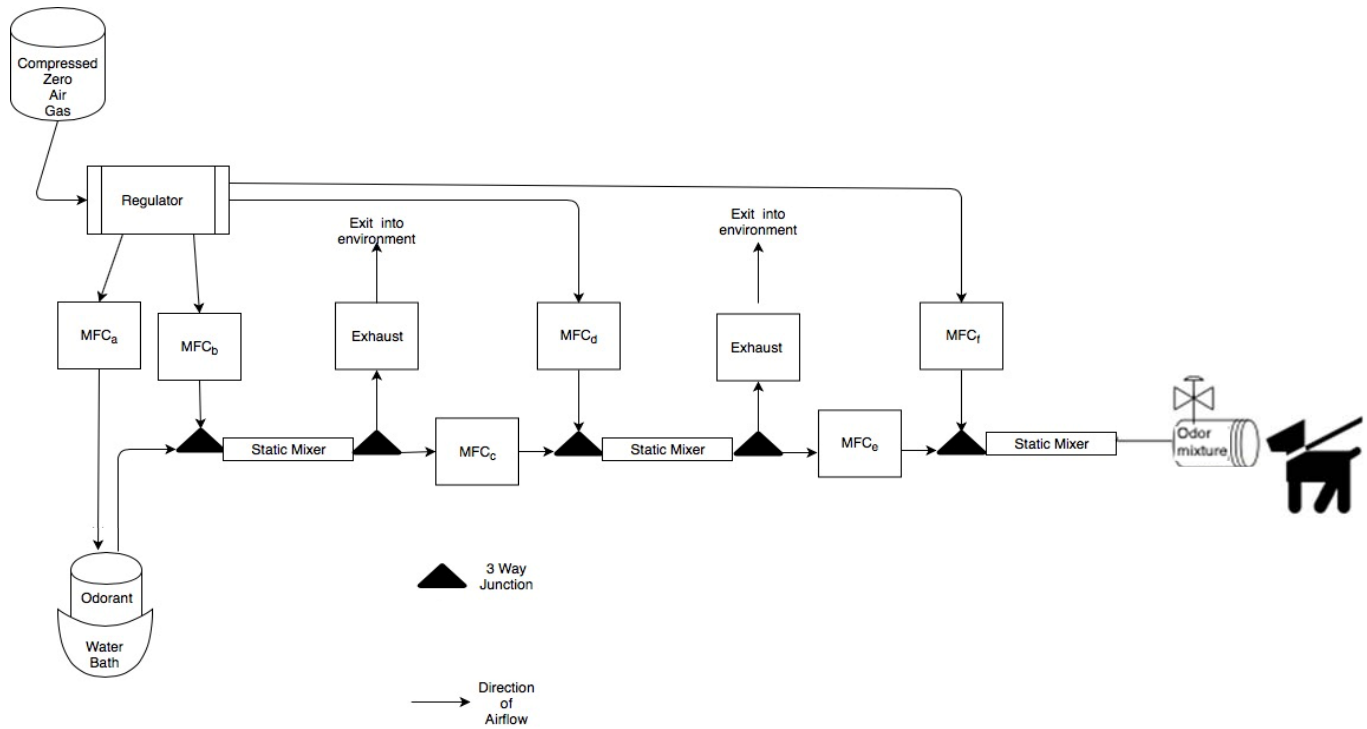

Figure 1. Schematic of the air-dilution olfactometer. Air flow path is shown from left to right and dog is shown evaluating the odor port. Briefly, compressed zero air is used to collect odorant from a sealed glass odorant jar in a water bath. The vapor stream is then systematically diluted using mass air flow controllers and turbulent static mixers. The odor line is the delivered to the odor port.

Figure 2. Calibration of Air Dilution Olfactometer. Left shows the systematic change in the Photo Ionization Detector output as a function of systematic changes in the air dilution olfactometer (R^2=0.998). X-axis shows the air dilution in effect on the olfactometer, and Y-axis shows the change in sensor output.

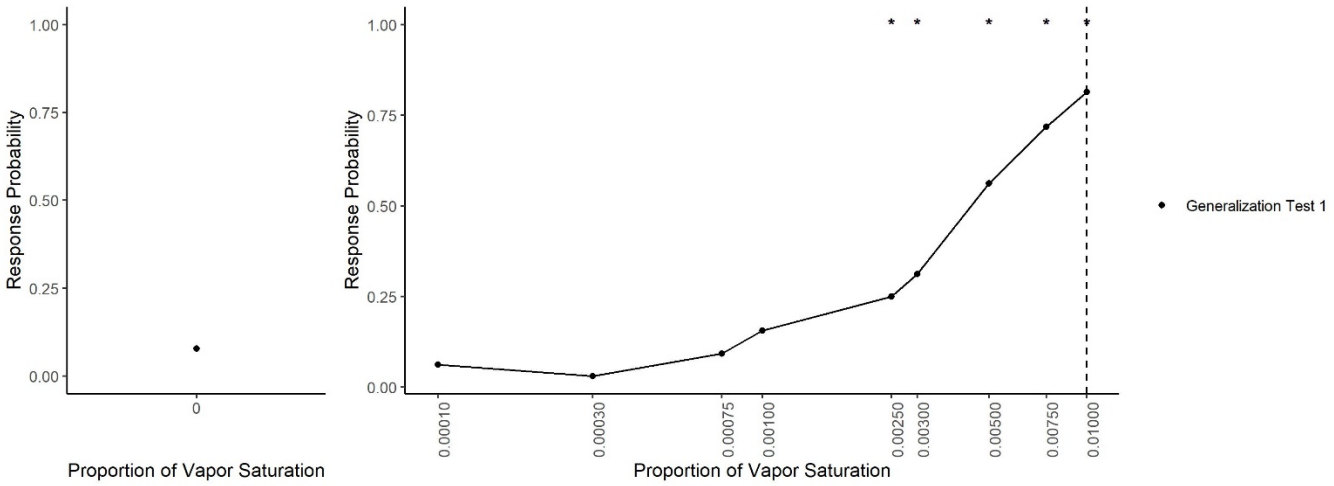

Figure 3. Generalization Test 1. Shows spontaneous generalization to a range of air dilutions of the trained odorant (0.01; dashed line). Left shows response to a clean air stimulus, Right shows the range of concentrations on a log scale. * indicates the probability of response was greater than the probability of response to the clean stimulus $p<0.05$.

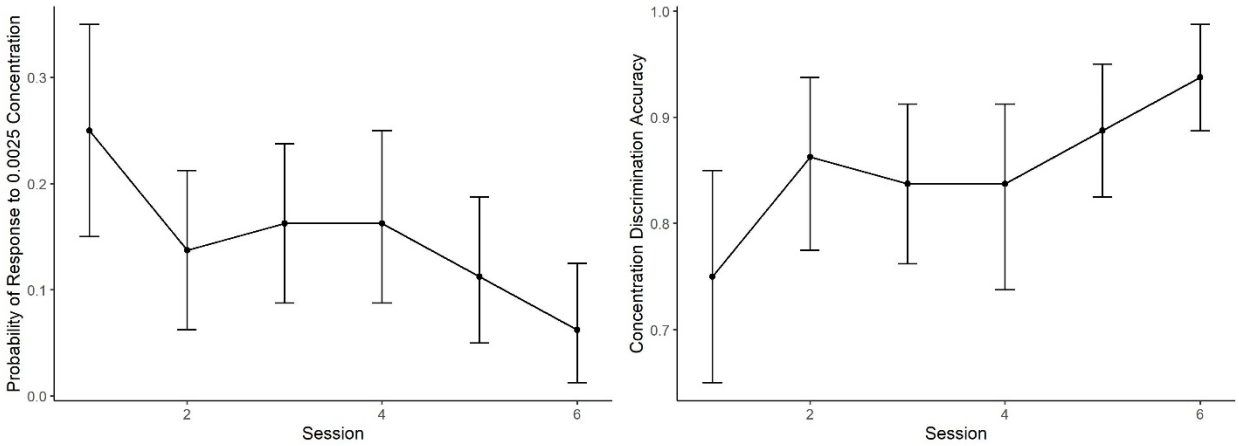

Figure 4. Concentration Discrimination Training. Shows mean performance and 95% confidence intervals during the concentration training sessions. Left shows the probability dogs alerted to the 0.0025 concentration, which decreased across sessions. Right shows overall session accuracy, where alerting to 0.01 was correct and not alerting to 0.0025 was correct.

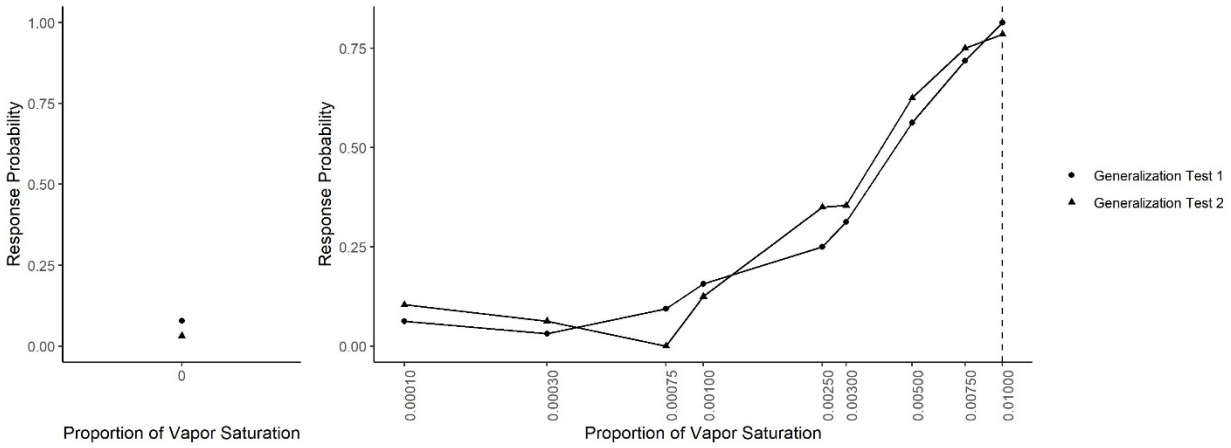

Figure 5. Combined generalization testing. Shows spontaneous generalization to a range of air dilutions of the trained odorant (0.01; dashed line) for Generalization Test 1 and Test 2. Left shows response to a clean air stimulus, Right shows the range of concentrations on a log scale.

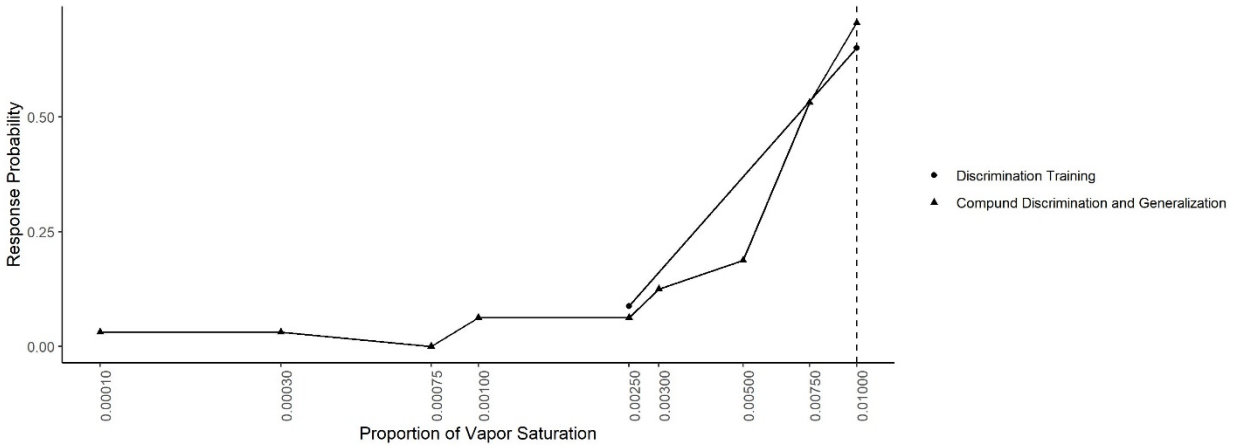

Figure 6. Concentration Discrimination and Compound Generalization. Shows the response probability for the last two sessions of the discrimination training phase (0.01 vs. 0.0025) and the Compound Discrimination and Generalization test. Dashed line shows the target odor concentration.

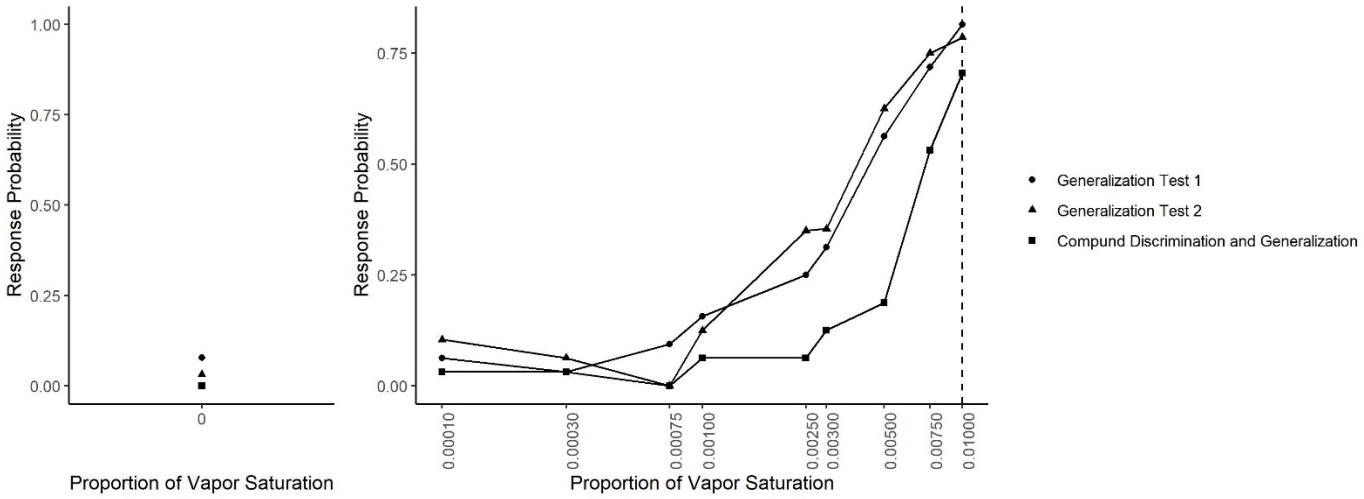

Figure 7. Generalization across all tests. Shows spontaneous generalization to a range of air dilutions of the trained odorant (0.01; dashed line) for Generalization Test 1, Test 2, and Compound Discrimination and Generalization. Left shows response to a clean air stimulus, Right shows the range of concentrations on a log scale.